LEVEL II

12

RADC-TR-79-147
Final Technical Report
June 1979

AD A072113

VALIDATION OF SOFTWARE RELIABILITY MODELS

Hughes Aircraft Company

R.E. Schafer J.F. Alter
J.E. Angus S.E. Emoto

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, New York 13441

79 07 31 013

This report has been reviewed by the RADC Information Office (OI) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RADC-TR-79-147 has been reviewed and is approved for publication.

APPROVED: Alan N. Sukert

ALAN N. SUKERT
Project Engineer

APPROVED: Wendall C Bauman

WENDALL C. BAUMAN, Col, USAF
Chief, Information Sciences Division

FOR THE COMMANDER: John P. Huss

JOHN P. HUSS
Acting Chief, Plans Office

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (ISIS), Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return this copy. Retain or destroy.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER RADC-TR-79-147	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) VALIDATION OF SOFTWARE RELIABILITY MODELS		5. TYPE OF REPORT & PERIOD COVERED Final Technical Report October 77 - November 78
		6. PERFORMING ORG. REPORT NUMBER FR-78-16-1263
7. AUTHOR(s) R.E. Schafer S.E. Emoto J.E. Angus J.F. Alter		8. CONTRACT OR GRANT NUMBER(s) F30602-77-C-0219
9. PERFORMING ORGANIZATION NAME AND ADDRESS Hughes Aircraft Company P.O. Box 3310 Fullerton CA 92634		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62702F 55812002
11. CONTROLLING OFFICE NAME AND ADDRESS Rome Air Development Center (ISIS) Griffiss AFB NY 13441		12. REPORT DATE June 1979
		13. NUMBER OF PAGES 180
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Same		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: Alan N. Sukert (ISIS)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Software reliability
Software debugging
Software reliability models
Software error models

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report presents the results of a study and investigation of software reliability models. In particular, the purpose was to investigate the statistical properties of selected software reliability models, including the statistical properties of the parameter estimates, and to investigate the goodness of fit of the models to actual software error data. The results indicate that the models fit poorly, generally due to in most part the vagaries of the data rather than shortcomings of the models.

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

172 350

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (Continued)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

EVALUATION

The increasing requirements for highly reliable software systems in such applications as command and control and avionics has led to the need for better specification of reliability requirements in software developments and for techniques for assessing the reliability of a delivered software system. This need has been expressed in numerous industry and government sponsored conferences, as well as documents such as the Joint Commanders' Software Reliability Working Group Report (November 1975). As a result, numerous efforts have been initiated to develop mathematical techniques for predicting the reliability of software systems. However, the application of these techniques has been delayed because they have not been subjected to sufficient validation efforts to check both model mathematics and model predictions.

This effort was initiated in response to this need for model validation and fits into the goals of RADC TPO No. 5, Software Cost Reduction in the subthrust of Software Quality (Software Modeling). This report summarizes the validation of several different mathematical models for predicting the reliability and error content of a software system against software error data extracted from actual Department of Defense software developments. This validation effort has investigated these models in terms of not only the validity of model assumptions and the properties of model parameter estimators, but also in terms of model predictability based upon statistical goodness-of-fit tests. The importance of this validation effort is that it provides the necessary statistical properties and model experimentation to permit the use of the investigated models on actual ongoing software developments.

The results of this model validation effort will lead to the use of these software reliability prediction models by Air Force software managers to accurately assess the status of their developments early enough to correct potential problems without increasing significantly the lift cycle costs. The results of this validation effort will also permit the specification of software reliability requirements on future software RFPs by insuring that the techniques do exist to verify that the delivered software meets these requirements. Finally, the models validated under this effort will be applicable to current software development projects, thus helping to increase the reliability and lower the life cycle cost of todays software systems.

Alan N. Sukert

ALAN N. SUKERT
Project Engineer

Section 0.0

SUMMARY

The purpose of this study was the validation of existing software (S/W) reliability models. This validation was accomplished by investigating the properties of model parameter estimates, by investigating the validity of the model internal assumptions, and by analyzing the goodness-of-fit of the models. These investigations were all made in terms of actual S/W error data from numerous (16) electronic system computer programs which represented a wide variety of system types.

The types of S/W reliability models studied were basically two: Poisson and Binomial models. The methods of parameter estimation investigated were also two: the maximum likelihood method and the least squares method.

The following results were obtained.

Methods of Estimation

- Both methods of estimation require numerical methods of solution for parameter estimation. Ordinarily, in this age of the digital computer, this would be no problem. However, definitive methods of determining the required starting points (for the numerical method) were not available and convergence problems were observed: either no convergence was obtained or convergence to absurd (e.g. negative values for parameters known to be positive) values.
- It was established that, with respect to precision, the MLEs were superior to the LSEs.
- The MLE method seemed to experience slightly more convergence problems than the LSE method.

Models

- The best fitting model was the most general (i.e., 3 parameters) of the Poisson models. A major advantage of these Poisson models is that they permit distinction between errors observed and errors corrected.

- It was clear that the fits were poor for all models. That is, poor fit was the uniformly common result.

- The poor fits were due to several causes:

  i) all of the model internal assumptions are not valid. For example, none of the models accommodate the introduction of errors in the debugging process.

  ii) the S/W debugging process is likely much more complex than the models assume.

  iii) the models seemed over-sensitive to the failure of data to "follow" the internal model assumptions.

Data Base

- Generally, the data sets failed to show a decreasing error rate as (calendar) time increases: even apart from random flucuations. This property of the data played havoc with the model fits. We suspect the data displayed this property because

  i) errors were introduced during the process of correcting discovered errors.

  ii) the man-power effort was not constant throughout the development process.

  iii) all of the S/W was not subject to test all of the time.

In short, while a great deal of usefil work has been accomplished in S/W reliability modeling, more work remains: mostly, we feel, in the area of S/W error data collection.

Section 1.0

INTRODUCTION

The objective of this effort was to validate selected existing S/W reliability models. In particular the specific objectives were

i) With regard to estimating the parameters in the models

- identify the MLEs and LSEs
- identify conditions under which the estimates do not exist
- investigate the statistical (small sample and large sample) properties of the estimates

ii) Validate internal assumptions of the model in terms of S/W development processes

iii) Test the goodness-of-fit of the models on actual S/W error data

iv) Develop methods of predicting the time necessary to remove a specified percent of remaining errors for each model

v) Compare efficiency of the various models.

We were successful in accomplishing all of these objectives except we were not able to work out the small sample exact estimator probability distributions nor were we able to obtain, except by simulation, other small sample properties; the problem being one of intractability of the analytical problems and the inability to find appropriate pivotal functions (functions of the parameters and the estimator(s) distributed free of the other parameters).

S/W reliability modeling is a tough, very tough, business in the following sense. First, no model, however attractive, will ever be accepted without verification with actual S/W error data; and rightfully so. Now the physical, stochastic process which generates the S/W error data is extremely complex and it is difficult to invent models with enough parameters to reflect the

process. The problem even goes deeper as illustrated by the following idea. It is entirely plausible that as S/W is debugged (and hence errors are discovered and corrected) errors are introduced. Now it is not impossible to modify the models (except for the IBM Model in a limited sense, none of the models we investigated accommodate the introduction of errors) to include a mechanism to allow for introducing errors. A most obvious, and simple, way is to include a probability, say p_I, that an error is introduced at each removal. But is such a parameter, namely p_I, estimable? That is, can it be estimated at all? It is unlikely that it can no matter how carefully the S/W data is "taken", for in order to estimate p_I one would have to have, in some form or other, the number of errors introduced at each removal and/or discovery. It is doubtful

i) that anyone knows that an error has been introduced or he wouldn't have introduced it

ii) that the error data as it occurs could be separated into those initially present and those introduced in correction/removal.

More to the point is the remark that there are more serious difficulties in validating S/W reliability models than not counting induced errors. The history of mathematical modeling is replete with examples of situations where the model "oversimplified" the process and yet was astonishingly successful in modeling the process. The problem with S/W reliability modeling is the data base; in particular the collection of data concerning concomitant variables. That is, there are variables which have measurable and important effects on the number of errors observed over any particular time. Such a variable is the number of man-hours per calendar hour devoted to the debugging process. This number is certainly rarely constant throughout the debugging process. There are other such variables; moreover, it is difficult to incorporate all of them in the various models; the models would contain too many parameters and the mathematics of the estimation process would become intractable. Thus the values assumed by these concomitant (not included in the model) variables must be recorded so that the data may be adjusted.

In any event the objectives previously mentioned were approached in terms of i) three (3) Poisson type S/W reliability models: Jelinski-Moranda (J-M), Shick-Wolverton (S-W) and a generalized Poisson model (GPM), ii) a binomial model and iii) an IBM non-homogeneous Poisson process type model which also arises in the context of (hardware) reliability growth.

The required S/W error data base was sixteen (16) data sets representing a variety of types of systems (i.e., of hardware systems).

The balance of this report describes the details and analyses involved in achieving the report objectives.

Section 2.0

DATA BASE

2.1 DATA BASE: GENERAL DESCRIPTION

The data base for this S/W reliability study was constructed from a total of 12 sources; three internal to Hughes, and nine from external sources. These sources represented a wide range of S/W applications including command and control, sonar, radar (ground fixed, ground mobile, shipboard), guidance and navigation, communications, and general information processing systems. A summary of this data is given in Table 2.1.1 along with the time period over which each set of data was collected, number of errors observed, language used, and approximate number of instructions.

2.2 DESCRIPTION OF DATA SETS AND PROCESSING RAW DATA

With the exception of the reference (II,2) data, all data for the S/W study was extracted from raw data which included such information as SPR or PTR number, routine suspected of having the error, date of error occurrence, an error category code, date of error fix, and various comments concerning the nature of the problem and how the fix was made. The data needed for the S/W models was the number of errors observed and removed in successive debugging-time intervals (the cut-off times for these time intervals were determined by dates when clustering of observations occurred). The number of errors and fixes and the length of any given debugging-time interval could not be extracted directly, however. The SPR's or PTR's corresponding to certain error category codes were deleted from the raw data because the error codes did not correspond to S/W failures. The types of "errors" described by these inadmissible codes include user requested changes, duplicate reports, documentation errors, hardware related reports, and questions (i.e., the SPR or PTR was used simply to make a statement or ask a question). Having removed these non-S/W failures it was then necessary in computing the debugging-time interval lengths to remove Saturdays, Sundays, and holidays since, in general, the debugging ceased on these days. If an SPR or PTR was opened or a fix made on a holiday or weekend, then that holiday or that entire weekend was counted in the debugging-time interval length.

TABLE 2.1.1. DESCRIPTION OF DATA SETS

Data Set Identification Number	Data Source	System Type	No. of Errors Observed	Time Period	No. of Instructions, Languages
1	Boeing	Avionics/Radar	2035	1973–1975	80,000 assembly language instructions and approx 40,000 lines of JOVIAL/J3B instructions (roughly equivalent to 240,000 assembly instructions)
2	Draper	Guidance, Navigation, and Control	7819	1967–1971	610,000 (Assembly language)
3	TRW (Project 2)	Command and Control	366	10/71–6/72	96,931 (JOVIAL J4)
4	TRW (Project 3) (Development)	Command and Control	4176	6/73–2/74	115,346 (JOVIAL J4)
5	TRW (Project 3) (Operational)	Command and Control	450	2/74–7/74	115,346 (JOVIAL J4)
6	TRW (Project 4)	Generalized Information Processing System	406	4/72–3/74	* (PWS)
7	IBM	Radar, Ground Fixed Command and Control Communications	3537	March 1, 1974–March 1, 1975	1,317,031 (CENTRAN)

*PWS macros not measurable in comparable units.

TABLE 2.1.1. DESCRIPTION OF DATA SETS (CONT)

Data Set Identification Number	Data Source (Project Name)	System Type	No. of Errors Observed	Time Period	No. of Instructions, Languages
8	Ref. (II,2)	**	1138	7/73–11/73	**
9	Ref. (II,2)	**	1483	7/73–11/73	**
10	Ref. (II,2)	**	2707	7/73–11/73	**
11	Ref. (II,2)	**	2362	7/73–11/73	**
12	Ref. (II,2)	**	26	**	**
13	Hughes (TPQ-37)	Radar, Ground Mobile	1789	11/72–11/77	20,000 (Assembly language)
14	Hughes (IPD/TAS)	Radar, Shipboard	503	6/77–2/78	32,200 (Assembly language)
†15	Hughes (SURTASS)	Sonar	447	10/76–10/77	32,420 (CMS2M)
15A		Sonar	318	10/76–10/77	17,046 (CMS2M)
15B		Sonar	25	10/76–10/77	5,967 (CMS2M)
15C		Sonar	261	10/76–10/77	9,407 (CMS2M)
16	Raytheon	Radar, Ground Fixed	1333	1/72–1/76	123,689 (JOVIAL J3)

**Unknown.

†15A, 15B, 15C represent distinct computer programs making up 15. The sum of the errors found in 15A, 15B, 15C do not add to 447 because they were adjusted for varying manpower assignments and percent of total software being tested.

2.3 FURTHER PROCESSING OF DATA

A common assumption in S/W reliability models is that all errors discovered in a given debugging-time interval are removed at the end of that interval. All raw data used in this study indicated that it was impossible to choose the debugging-time intervals in this way because of the way find-dates and fix-dates were clustered. In view of this difficulty, three types of data sets were constructed from the raw data (see Figure 2.2.1). These consisted of the number of "finds" for successive debugging-time intervals, the number of "fixes" for successive debugging intervals, and a combination of these whereby a column of data with the cumulative number of fixes made as a function of debugging-time was combined with the "find" data (examples of these are shown in Table 2.2.1). The models can be fit to the "find" data by assuming that the errors are actually removed at the end of each debugging-time interval and that the "fix" dates are inaccurate. Alternately, the models can be fit to the "fix" data assuming that the number of finds in a particular debugging-time interval between successive clusters of fixes is equal to the number of fixes in that time interval. Finally, when possible, the models can utilize the combined find/fix data which result in a different number of "fixes" at the end of a debugging-time interval than "finds" within that interval. (These three data types are easily utilized by most of the models studied here by defining the cumulative number of errors removed by the end of each debugging-time interval as a known parameter instead of as the cumulative sums of the sequence of observations of errors which are random quantities. See Section 3.0 for further discussion.)

As a final data preprocessing effort, each "find," "fix," and combined "find" and "fix" data set was converted so that within each debugging-time interval, at least 10 errors were observed. This was done by combining successive time intervals until the total number of errors observed exceeded or equaled 10. This effort was initiated to insure the validity of the limiting distribution of the statistic used for testing goodness-of-fit (see Section 5.2).

Table 2.2.2 shows the numbering conventions used for each data set.

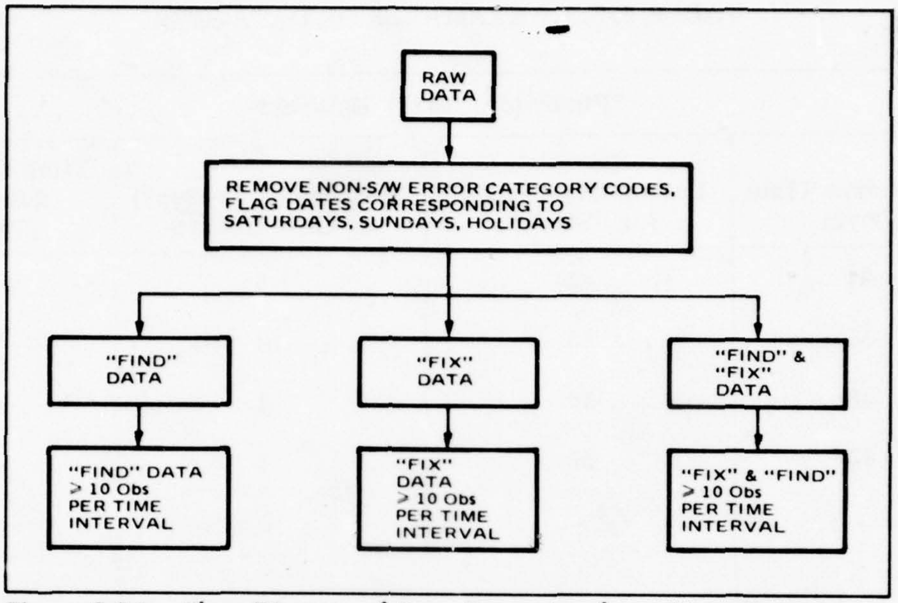

Figure 2.2.1. Flow Diagram of Data Processing for a Given Data Source

TABLE 2.2.1. EXAMPLES OF DATA SETS

"Find" (or "Fix") Data Set

Cumulative Time (Days)	Cumulative "Finds" (or "Fixes")	"Finds" (or "Fixes") in Time Interval	Time Interval Length (Days)
34	5	5	34
35	15	10	1
39	16	1	4
43	20	4	4
.	.	.	.
.	.	.	.
.	.	.	.

Combined "Find" and "Fix" Data Set

Cumulative Time (Days)	Cumulative "Finds"	"Finds" in Time Interval	Time Interval Length (Days)	Cumulative "Fixes"
2	1	1	2	1
3	3	2	1	1
15	4	1	12	2
29	5	1	14	3
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

TABLE 2.2.2. DATA SET NUMBERING CONVENTION

15A – fdG

Data Source Designator (See Table 2.1.1)

Data Type Designator:

ff – Combined "find" and "fix"

ffG – Combined "find" and "fix" grouped for $\geq$10 observations per time interval

fd – "Find"

fdG – "Find" grouped for $\geq$10 observations per time interval

fx – "Fix"

fxG – "Fix" grouped for $\geq$10 observations per time interval

Section 3.0

DESCRIPTION OF MODELS

3.1 GENERALIZED POISSON MODEL

The generalized Poisson model (GPM) is a generalization of the S-W and J-M models discussed in references (III, 9), (III, 10), (III, 20), and (III, 24). The GPM assumes that the number of errors N_i observed in the ith debugging-time interval τ_i has a Poisson distribution with mean value

$$E(N_i) = \phi (N-M_{i-1}) \tau_i^{\alpha}$$

where ϕ is a constant of proportionality, N is the total number of errors present in the S/W initially, and M_j is the total number of errors removed up to the end of the jth debugging-time interval. It is assumed that when errors are removed, they are removed at the ends of the debugging-time intervals. Some authors have assumed that

$$M_j = \sum_{i=1}^{j} N_j ;$$

i.e., that all errors found in each debugging-time interval are removed at the end of that interval. This is an unnecessary assumption and practically is never satisfied. We may treat $M_0 \equiv 0$, M_1, M_2, ..., M_{k-1} as known constants and assume that successive observations N_1, N_2, ..., N_k are independent. The joint probability function of N_1, ..., N_k is

$$P \left| N_1 = n_1, \ldots, N_k = n_k \right| = \prod_{i=1}^{k} \frac{\left| \phi (N-M_{i-1}) \tau_i^{\alpha} \right|^{n_i} e^{-\phi (N-M_{i-1}) \tau_i^{\alpha}}}{n_i !}$$

It should be pointed out that if $\tau_1 = \tau_2 = \ldots = \tau_k$, then the parameter α (α describes the degree of dependence of the expected number of errors on the length of the time interval) is not identifiable, i.e., τ_i^{α} may be absorbed into the constant ϕ and the model reduces to a two (N, ϕ) parameter model. Because of the assumptions

$$\mathrm{Var}\,(N_i) = \phi\,(N-M_{i-1})\,\tau_i^{\alpha}$$

$$\mathrm{Cov}\,(N_i, N_j) = 0, \; i \neq j.$$

For the case when $\alpha = 1$, this model becomes the J-M model and when $\alpha = 2$, it becomes the S-W model. In general, α should be positive since the number of errors, on the average, observed in a time interval should increase as the time interval length increases, and should decrease to 0 as the time interval length shrinks to 0.

3.2 IBM MODEL

The IBM model is based on the non-homogeneous Poisson process (see (II, 2). The basic assumptions concerning the non-homogeneous Poisson process, say N(t), are:

1. $N(t)$, $t \geq 0$, is an integer valued random variable with $N(0) = 0$ almost surely.

2. For $0 \leq t_1 < t_2 < \ldots < t_k$, the random variables $N(t_1)$, $N(t_2) - N(t_1)$, $N(t_4) - N(t_3)$, ..., $N(t_k) - N(t_{k-1})$ are stochastically independent.

3. $P\{N(t+h) - N(t) \geq 2\} = o(h)$* as $h \to 0+$

4. $P\{N(t+h) - N(t) = 1\} = \lambda(t)h + o(h)$ as $h \to 0+$ where $\lambda(t)$ is integrable over $[0, T]$ for all T, $\lambda(t) \geq 0$ for all $t \geq 0$.

From these assumptions the probability distribution of $N(t) - N(t_0)$ can be derived for any $t > t_0 \geq 0$.

Let $P_n(t) = P\{N(t) - N(t_0) = n\}$, $n = 0, 1, \ldots$. Then,

$$P_0(t+h) = P_0(t)\,(1 - \lambda(t)h) + o(h) \text{ as } h \to 0+.$$

Rearranging and dividing by $h > 0$ and letting $h \to 0$,

$\dfrac{dP_0(t)}{dt} = -\lambda(t)P_0(t)$ with the initial condition $P_0(t_0) = 1$. Solving this linear differential equation yields

$$P_0(t) = \exp\left(-\int \lambda(t)\,dt + C\right)$$

*$f(h) = o(h)$ as $h \to 0+$ means $f(h)/h \to 0$ as $h \to 0$ and $h > 0$.

and utilizing the initial condition gives

$$P_0(t) = \exp\left(-\int_{t_0}^{t} \lambda(s)\, ds\right).$$

Letting $m(t) = \int_0^t \lambda(s)\, ds$, we see that

$$P_0(t) = \exp(-(m(t) - m(t_0))).$$

For $n \geq 1$, we have the difference equation

$$P_n(t+h) = P_n(t)(1-h\lambda(t)) + P_{n-1}(t)\, h\lambda(t) + o(h) \text{ as } h \to 0+$$

which leads to the differential equation

$$\frac{dP_n(t)}{dt} = -\lambda(t)\, P_n(t) + \lambda(t)\, P_{n-1}(t), \quad n = 1, 2, \ldots.$$

This system of differential equations may be solved recursively, having found $P_0(t)$. E.g.,

$$\begin{aligned} \frac{dP_1(t)}{dt} &= -\lambda(t)\, P_1(t) + \lambda(t)\, P_0(t) \\ &= -\lambda(t)\, P_1(t) + \lambda(t) \exp(-(m(t) - m(t_0))) \end{aligned}$$

with initial condition $P_1(t_0) = 0$. This differential equation leads to

$$P_1(t) = (m(t) - m(t_0)) \exp(-(m(t) - m(t_0))).$$

Continuing in this fashion gives

$$\begin{aligned} P_n(t) &= P\{N(t) - N(t_0) = n\} \\ &= \frac{[m(t) - m(t_0)]^n}{n!}\, e^{-(m(t) - m(t_0))} \end{aligned}$$

for $n = 0, 1, 2, \ldots$.

For software reliability modeling, $N(t)$ can be taken to be the total number of errors detected in time $(0, t)$. The expected number of error occurrences in $(0, t)$, $E(N(t)) = m(t)$, is assumed to satisfy the following conditions:

$$m(t+h) = m(t) + bh(a - m(t)) + o(h) \text{ as } h \to 0+,$$

where $b > 0$ is a constant of proportionality, and $a > 0$ is the expected number of errors to be found in infinite time, i.e.,

$$a = \lim_{t \to \infty} m(t)$$

(We have required, of course, that a be finite.) That is, the expected number of errors in time $[0, t+h]$ is approximately the expected number of errors in time $(0, t)$ plus an amount proportional to h and the expected number of errors observable in time (t, ∞) (analogous to the number of errors remaining, $N-M_j$, for the generalized Poisson model). Dividing the difference equation for the mean value function by $h > 0$ and letting $h \to 0$ we obtain

$$\frac{dm(t)}{dt} = b\,(a-m(t))$$

with initial condition $m(0) = 0$. The solution is, of course,

$$m(t) = a\,(1-e^{-bt}),\ t \geq 0.$$

To summarize these considerations, we have for any $t > t_0 > 0$,

$$P\{N(t) - N(t_0) = n\} = \frac{\{a(e^{-bt_0} - e^{-bt})\}^n}{n!}\, e^{-a(e^{-bt_0} - e^{-bt})}$$

and for $t_0 = 0$,

$$P\{N(t) = n\} = \frac{\{a(1-e^{-bt})\}^n}{n!}\, e^{-a\,(1-e^{-bt})}$$

For $0 \leq t_1 < t_2 < \ldots < t_k$, the joint probability function

$$P\{N(t_1) = n_1, \ldots, N(t_k) = n_k\}$$

is difficult to derive. However, the joint probability function of the increments

$$Z_1 = N(t_1),\ Z_2 = N(t_2) - N(t_1),\ \ldots,\ Z_k = N(t_k) - N(t_{k-1})$$

is straight-forward since the increments are independent. We have

$$P\{Z_1 = z_1, \ldots, Z_k = z_k\} = \prod_{i=1}^{k} \frac{\left[a\left(e^{-bt_{i-1}} - e^{-bt_i}\right)\right]^{z_i}}{z_i\,!}\, e^{-a\left(e^{-bt_{i-1}} - e^{-bt_i}\right)}$$

where $t_0 \equiv 0$.

Since for $t > t_0 \geq 0$, $N(t) - N(t_0)$ has the Poisson distribution shown above,

$$E\,(N(t) - N(t_0)) = a\left(e^{-bt_0} - e^{-bt}\right)$$

$$\mathrm{Var}\,(N(t) - N(t_0)) = a\left(e^{-bt_0} - e^{-bt}\right).$$

In fitting the IBM model to data, the data needed are the cumulative times $0 \leq t_1 < t_2 < \ldots < t_k$ and cumulative errors $N(t_1) < N(t_2) < \ldots < N(t_k)$ observed. The cumulative errors are converted to increments Z_i since the LSEs and MSEs are derived in terms of the increments rather than the cumulative errors. This is somewhat different than what was done in (II,2). There, the LSE's of a and b were derived by minimizing

$$S(a,b) = \sum_{i=1}^{k} \left(N(t_i) - a\left(1-e^{-bt_i}\right)\right)^2.$$

However, when testing goodness-of-fit, it will be shown (see Section 5.2) that utilizing the increments Z_i rather than the cumulative errors in fitting the models makes complete use of the data whereas using cumulative errors makes use of only one data point out of the k observations.

In the discussion of Section 5.2, it will be necessary to know the covariance matrix for the normalized variables

$$Y_1 = \frac{N(t_1) - m(t_1)}{\sqrt{m(t_1)}}, \ldots, Y_k = \frac{N(t_k) - m(t_k)}{\sqrt{m(t_k)}},$$

i.e., $\mathrm{cov}\,(Y_1, \ldots, Y_k) = \sum$

where

$$\sum_{ij} = E(Y_i\, Y_j),\ i = 1, \ldots, k,\ j = 1, \ldots, k.$$

For

$$i = j, \sum_{ij} = \sum_{ii} = E(Y_i^2) = 1.$$

For $i > j$, we have

$$\sum_{ij} = E(Y_i Y_j) = E\left\{\frac{(N(t_i) - m(t_i))}{\sqrt{m(t_i)}} \cdot \frac{(N(t_j) - m(t_j))}{\sqrt{m(t_j)}}\right\}$$

$$= E\left\{\left[\frac{(N(t_i) - m(t_i))}{\sqrt{m(t_i)}} - \frac{(N(t_j) - m(t_j))}{\sqrt{m(t_j)}}\right] \cdot \frac{(N(t_j) - m(t_j))}{\sqrt{m(t_j)}} + \frac{(N(t_j) - m(t_j))^2}{m(t_j)}\right\}$$

$$= E\left[(Y_i - Y_j)\, Y_j\right] + E(Y_j^2) = E(Y_j^2) = 1$$

Since $Y_i - Y_j$ and Y_j are independent. Thus, the covariance matrix is simply a $k \times k$ matrix of 1's:

$$\sum = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

As mentioned earlier, the constant "a" is analogous to the constant N in the GPM. There is a very important difference, though. N is the number of errors initially in the software for the GPM whereas a is the expected number of errors which could be observed in unlimited time, i.e., if we let $N(\infty)$ denote $\lim_{t \to \infty} N(t)$ (a well defined random variable), $E(N(\infty)) = a$ since $N(\infty)$ has a Poisson distribution with mean a. Since $N(\infty)$ is not constant almost surely, one can interpret the IBM model as one in which errors are allowed to be introduced during debugging.

As a final observation on the IBM model, it should be pointed out that when the data actually behaves as though they were generated from a homogeneous Poisson process, i.e., $m(t) = \lambda t$ for some positive constant λ, the parameter estimators will reflect this by the estimate of b decreasing towards zero while the estimate of a increases without bound, with the product ab remaining constant. That is, letting $ab = \lambda$ or $a = \lambda/b$, we see that

$$\lim_{b \to 0} m(t) = \lim_{b \to 0} a(1-e^{-bt}) = \lim_{b \to 0} \frac{\lambda}{b}(1-e^{-bt})$$

$$= \lim_{b \to 0} \frac{\lambda}{b}(1 - (1 - bt + o(b))) = \lambda t,$$

the mean value function for a homogeneous Poisson process. This happened frequently as shown in Tables 5.1.1 through 5.1.16.

3.3 BINOMIAL MODEL

The binomial models discussed here are based on a model suggested by P. A. P. Moran in (II,5). Basically, it is assumed that there are N errors present initially. Errors are then observed in successive time intervals $\tau_1, \tau_2, \ldots, \tau_k$ under the assumption that at the end of each time interval, the errors observed during that time interval are removed from the S/W, thus decreasing the total number of errors in the S/W by that number. As with the GPM (and the S-W and J-M models) no consideration is given to the possibility of creating new errors in the process of removing errors. The joint probability function is derived under the assumption that N_i, the number of errors observed in τ_i, has a binomial distribution with the n-parameter equal to the total errors remaining in the S/W, and the p-parameter equal to some function of τ_i and possibly an unknown constant. This function must satisfy the following assumptions (denoting the function by $p(\tau_i)$):

1. $0 \le p(\tau_i) \le 1$ for all $\tau_i \ge 0$

2. $p(\tau_i)$ is an increasing function of $\tau_i \ge 0$

3. $p(0) = 0, \quad \lim_{\tau_i \to \infty} p(\tau_i) = 1.$

Since $p(\tau_i)$ is interpreted as a probability of detection in time τ_i, assumption 1 is necessary while assumptions 2 and 3 mean that given no time, no errors can be detected while given unlimited time, an error will certainly be detected.

Letting $p_i = p(\tau_i)$ we may now derive the joint probability function for $N_1, \ldots, N_k$. Clearly, the probability function for N_1 is

$$P\{N_1 = n_1\} = \binom{N}{n_1} p_1^{n_1} (1-p_1)^{N-n_1}, \; n_1 = 0, \ldots, N.$$

Given N_1, the probability function for N_2 is

$$P\{N_2 = n_2 \mid N_1 = n_1\} = \binom{N-n_1}{n_2} p_2^{n_2} (1-p_2)^{N-n_1-n_2},$$

for $n_2 = 0, 1, \ldots, N-n_1$ so that the joint probability function of N_1 and N_2 is

$$P\{N_1 = n_1, N_2 = n_2\} = \binom{N}{n_1}\binom{N-n_1}{n_2} p_1^{n_1} p_2^{n_2} (1-p_1)^{N-n_1} (1-p_2)^{N-n_1-n_2}$$

for $n_1 = 0, 1, \ldots, N, n_2 = 0, 1, \ldots, N-n_1$. Continuing in this fashion, we get

$$P\{N_1 = n_1, \ldots, N_k = n_k\} =$$

$$\binom{N}{n_1}\binom{N-n_1}{n_2}\cdots\binom{N - \sum_{j=1}^{k-1} n_j}{n_k} p_1^{n_1} \ldots p_k^{n_k} (1-p_1)^{N-n_1} \ldots (1-p_k)^{N - \sum_{j=1}^{k} n_j}$$

$$= \frac{N!}{n_1!n_2! \ldots n_k!\left(N - \sum_{j=1}^{k} n_j\right)!} p_1^{n_1} \ldots p_k^{n_k} (1-p_1)^{N-n_1} \ldots (1-p_k)^{N - \sum_{j=1}^{k} n_j},$$

for

$$\begin{cases} n_1 = 0, 1, \ldots, N \\ n_2 = 0, 1, \ldots, N-n_1 \\ \cdot \\ \cdot \\ \cdot \\ n_k = 0, 1, \ldots, N - \sum_{j=1}^{k-1} n_j \end{cases}$$

We have the following expectation values:

$$E\left(N_j \mid N_1 = n_1, \ldots, N_{j-1} = n_{j-1}\right) = \left(N - \sum_{i=1}^{j-1} n_i\right) p_j$$

$$\mathrm{Var}\left(N_j \mid N_1 = n_1, \ldots, N_{j-1} = n_{j-1}\right) = \left(N - \sum_{i=1}^{j-1} n_i\right) p_j (1-p_j).$$

Two different functions were considered for $p(\tau_i)$. The first was

$$p(\tau_i) = \tau_i/(\tau_i+c), \ c > 0$$

which corresponds to Binomial I in subsequent sections. This function clearly satisfies the assumptions 1, 2, and 3 and arises as the probability distribution of the ratio of two suitably chosen independent exponential random variables.

The second choice was

$$p(\tau_i) = 1-e^{-a\tau_i}, \; a > 0.$$

This function clearly satisfies assumptions 1, 2, and 3 and corresponds to Binomial II in subsequent sections. This function is recognized as the exponential distribution function.

The main rationale for choosing these functions was the presence of only one unknown parameter along with ease of computation using these functions.

The binomial models are not appropriate for use with the combined "fix" and "find" data since the errors observed in each time interval are assumed removed at the end of their respective time intervals. Also, the development of MLE's for unknown parameters N and a or c is very complicated and costly due to the difficulties involved with differentiating combinatorials and factorials. It is possible to introduce approximations (e.g., Stirling's approximation for large factorials) as in (II,5) but the conditions under which these would be valid could not possibly apply in all applications. Hence, only LSEs were derived in this study for the binomial models by minimizing

$$\sum_{i=1}^{h} (N_i - E(N_i | N_1, \ldots, N_{i-1}))^2$$

with respect to the unknown parameters N and a or c (see Section 4.1.4).

Section 4.0

METHODS OF ESTIMATION

4.1 TECHNIQUES OF ESTIMATING MODEL PARAMETERS

4.1.1 Basic Statistical Techniques

Two different statistical methods were used to estimate the parameters of the models of Section 3. These methods were designed to yield a set of parameter estimates for a data set. Among these sets, the most often estimated parameter was N, the unknown total number of errors present. In the IBM growth model of Section 3.2, the estimate of the parameter a represents the total number of errors expected, and includes errors added through performing corrections. The other parameters estimated in each set relate to the shaping and scaling of the time to error distributions.

The basic methods employed were those of least squares estimators (LSEs) and maximum likelihood estimators (MLEs).

The LSEs of a parameter (or set of parameters), say p, are those values, $\tilde{p}$, which minimize the sum of the squares of the difference between the observed values of a random variable, X, and expectations, E(X). I. e.

$$S = \sum_{i=1}^{k} (X_i - E(X_i))^2$$

must be minimized where the value $E(X_i)$ is dependent upon the value of p.

For the maximum likelihood method of obtaining estimators, the probability density function, f(X), is needed. The likelihood function L is the joint probability function of the k observations of the random variable X. The MLE's are those values $\hat{p}$ of the parameters which jointly maximize the function L for a fixed set of k observations. The values for $\hat{p}$ must be in the allowable parameter space; e.g. an MLE for the variance of the random variable must be non-negative.

For observations of the random variables considered here, the probability density function and the square $(X-E(X))^2$ (which depends upon the parameter p) are both

positive. Thus both logarithm L and logarithm S are well defined functions. Since the logarithmic function is monotone increasing, the maximum values of L and $\ell n L$ (or S and ℓn S) occur at the same parameter values $\hat{p}$ (or $\tilde{p}$). So, to simplify calculations, ℓn L (or ℓn S), rather than the likelihood function L itself (or sum of squares S), is maximized. In particular, when L is a product,

$$L = \prod_{i=1}^{k} f(X_i), \quad \ell n L = \ell n \prod_{i=1}^{k} f(X_i) = \sum_{i=1}^{k} \ell n f(X_i).$$

This follows since the logarithm of a product equals the sum of the logarithms, and is quite useful to actual calculations.

4.1.2 Numerical Methods of Solution

Due to the multiplicity of the parameters in the models, simultaneous equations must be solved in order to find the MLEs and LSEs. The Newton-Raphson method of finding roots to simultaneous equations (in some cases the relevant first partial derivatives equated to zero) was used frequently. This is an iterative process which uses tangent lines as approximations to the curve $y = F(X)$ in order to seek out the roots of the equation $F(X) = 0$. The derivative, $F'(X)$, of the function is used in choosing the next approximation to the root itself through the recursive relation

$$X_{i+1} = X_i + \frac{F(X_i)}{F'(X_i)} .$$

An error bound $\epsilon > 0$ is needed to terminate the process when $|X_{i+1} - X_i| < \epsilon$. The root is then approximated to be X_{i+1}, the value to which the successive approximations have converged. Values of ϵ on the order of 10^{-8} to 10^{-6} were used in this study.

4.1.3 Estimation for the Poisson Models

Computerized methods of obtaining estimators for the parameters of the "Poisson" type models described in Section 3.1 were used. In this case it was assumed that each N_i (number of errors observed in the ith time interval) had a Poisson distribution with mean

$$\lambda = \phi(N-M_{i-1}) \tau_i^{\alpha},$$

with the two particular cases of $\alpha = 1$ (J-M) and $\alpha = 2$ (S-W).

Here the LSEs are values $\tilde{N}$ and $\tilde{\phi}$ (and $\tilde{\alpha}$ in the three parameter case) which minimize the sum

$$S = \sum_{i=1}^{k} \left[N_i - \phi(N-M_{i-1}) \tau_i^{\alpha} \right]^2 .$$

The LSEs are the solutions to the equations obtained by setting the first partial derivatives of S (with respect to N, ϕ, and α) equal to zero. In the two parameter case this results in the least squares equations:

$$\frac{\partial S}{\partial N} = \frac{\partial S}{\partial \phi} = 0$$

where

$$\frac{\partial S}{\partial N} = -2\phi \sum_{i=1}^{k} N_i \tau_i^{\alpha} + 2\phi^2 \sum_{i=1}^{k} (N-M_{i-1}) \tau_i^{2\alpha}$$

$$\frac{\partial S}{\partial \phi} = -2 \sum_{i=1}^{k} N_i (N-M_{i-1}) \tau_i^{\alpha} + 2\phi \sum_{i=1}^{k} (N-M_{i-1})^2 \tau_i^{2\alpha}$$

and $\alpha = 1$ or 2.

In the three parameter model, the least squares equations are:

$$\frac{\partial S}{\partial N} = \frac{\partial S}{\partial \phi} = \frac{\partial S}{\partial \alpha} = 0$$

where

$$\frac{\partial S}{\partial N} = -2\phi \sum_{i=1}^{k} N_i \tau_i^{\alpha} + 2\phi^2 \sum_{i=1}^{k} (N-M_{i-1}) \tau_i^{2\alpha}$$

$$\frac{\partial S}{\partial \phi} = -2 \sum_{i=1}^{k} N_i (N-M_{i-1}) \tau_i^{\alpha} + 2\phi \sum_{i=1}^{k} (N-M_{i-1})^2 \tau_i^{2\alpha}$$

$$\frac{\partial S}{\partial \alpha} = -2\phi \sum_{i=1}^{k} N_i(N-M_{i-1})\tau_i^{\alpha} \ln \tau_i + 2\phi^2 \sum_{i=1}^{k} (N-M_{i-1})^2 \tau_i^{2\alpha} \ln \tau_i.$$

The iterative procedure for solving these equations requires starting points for N (and α in the three parameter case), with the starting point for N being

$$\geq \sum_{i=1}^{k} N_i$$

(the total number of errors already observed).

In the two parameter cases (i. e. α known) ϕ was eliminated from the least squares equations leaving one equation to solve for $\tilde{N}$ using the Newton-Raphson method. $\tilde{\phi}$ was then computed using $\tilde{N}$ and one of the original least squares equations. For the three parameter case the parameter ϕ was eliminated from the least squares equations and a two-dimensional Newton-Raphson method was used to find $\tilde{N}$ and $\tilde{\alpha}$ (having $\tilde{N}$ and $\tilde{\alpha}$, $\tilde{\phi}$ was easily computed using one of the original least squares equations). It was assumed that the estimate of N need not be an integer (although $\tilde{N}$ must satisfy)

$$\tilde{N} \geq \sum_{i=1}^{k} N_i).$$

This adjustment in assumptions was necessary to facilitate computations. Nearest integer values of $\tilde{N}$ are adopted in tabulating obtained results. This convention was used throughout the models.

The MLEs are those estimates $\hat{N}$ and $\hat{\phi}$ (and $\hat{\alpha}$ in the three parameter case) which maximize the logarithm of the likelihood function,

$$\ln L = \sum_{i=1}^{k} \ln \left\{ \left[\phi(N-M_{i-1})\tau_i^{\alpha} \right]^{N_i} e^{-\phi(N-M_{i-1})\tau_i^{\alpha}} \Big/ N_i! \right\}.$$

In particular, in the two parameter case, the MLE's are the solutions to the likelihood equations:

$$\frac{\partial \ln L}{\partial N} = \frac{\partial \ln L}{\partial \phi} = 0$$

where

$$\frac{\partial \ell n\, L}{\partial N} = \sum_{i=1}^{k} \frac{N_i}{N-M_{i-1}} - \phi \sum_{i=1}^{k} \tau_i^{\alpha}$$

$$\frac{\partial \ell n\, L}{\partial \phi} = \frac{1}{\phi} \sum_{i=1}^{k} N_i - \sum_{i=1}^{k} (N-M_{i-1}) \tau_i^{\alpha}$$

and $\alpha = 1$ or 2.

In the three parameter case, the MLE's were the solutions to the likelihood equations:

$$\frac{\partial \ell n\, L}{\partial N} = \frac{\partial \ell n\, L}{\partial \phi} = \frac{\partial \ell n\, L}{\partial \alpha} = 0$$

where

$$\frac{\partial \ell n\, L}{\partial N} = \sum_{i=1}^{k} \frac{N_i}{N-M_{i-1}} - \phi \sum_{i=1}^{k} \tau_i^{\alpha}$$

$$\frac{\partial \ell n\, L}{\partial \phi} = \frac{1}{\phi} \sum_{i=1}^{k} N_i - \sum_{i=1}^{k} (N-M_{i-1}) \tau_i^{\alpha}$$

$$\frac{\partial \ell n\, L}{\partial \alpha} = \sum_{i=1}^{k} N_i\, \ell n\, \tau_i - \phi \sum_{i=1}^{k} (N-M_{i-1}) \tau_i^{\alpha}\, \ell n\, \tau_i .$$

The iterative procedure for solving these equations requires starting points for N (and α in the three parameter case), with the starting point for N being

$$\geq \sum_{i=1}^{k} N_i .$$

As with the LS method, in the two parameter case ϕ was eliminated from the likelihood equations leaving one equation to solve for $\hat{N}$ using the Newton-Raphson method. $\hat{\phi}$ was then computed using $\hat{N}$ and one of the original likelihood equations. In the three parameter case ϕ was again eliminated from the likelihood equations, leaving only two equations to solve for $\hat{N}$ and $\hat{\alpha}$ using a two-dimensional Newton-Raphson method ($\hat{\phi}$ was computed using $\hat{N}$, $\hat{\alpha}$ and one of the original likelihood equations).

4.1.4 Estimation for the Binomial Models

For the binomial models described in Section 3.3, where the number of observations, N_i, is assumed to have a binomial distribution with parameters p_i and $N-M_{i-1}$, two separate cases were considered:

i) $$p_i = \frac{\tau_i}{\tau_i + c}$$

ii) $$p_i = 1-e^{-a\tau_i}$$

with both c and a assumed positive.

Here the least squares method finds those estimates $\tilde{N}$ and $\tilde{c}$ (or $\tilde{a}$) which minimize the sum

$$S = \sum_{i=1}^{k} \left[N_i - (N-M_{i-1})p_i \right]^2,$$

where

$$M_{i-1} = \sum_{j=1}^{i-1} N_j, \; i=2, \ldots, k, \; M_0 \equiv 0.$$

Specifically, for the binomial models, the LSEs are the solutions to the equations:

$$\frac{\partial S}{\partial N} = \frac{\partial S}{\partial c} = 0 \quad \text{or} \quad \frac{\partial S}{\partial N} = \frac{\partial S}{\partial a} = 0.$$

For the model having

$$p_i = \frac{\tau_i}{\tau_i + c},$$

the partial derivatives are given by:

$$\frac{\partial S}{\partial N} = -2 \sum_{i=1}^{k} N_i \frac{\tau_i}{\tau_i + c} + 2 \sum_{i=1}^{k} (N-M_{i-1}) \left(\frac{\tau_i}{\tau_i + c}\right)^2$$

$$\frac{\partial S}{\partial c} = 2 \sum_{i=1}^{k} N_i(N-M_{i-1}) \frac{\tau_i}{(\tau_i + c)^2} - 2 \sum_{i=1}^{k} (N-M_{i-1})^2 \frac{\tau_i^2}{(\tau_i + c)^3} .$$

For the model having

$$p_i = 1-e^{-a\tau_i} ,$$

$$\frac{\partial S}{\partial N} = -2 \sum_{i=1}^{k} N_i (1-e^{-a\tau_i}) + 2 \sum_{i=1}^{k} (N-M_{i-1}) (1-e^{-a\tau_i})^2$$

$$\frac{\partial S}{\partial a} = -2 \sum_{i=1}^{k} N_i\tau_i(N-M_{i-1}) e^{-a\tau_i} + 2 \sum_{i=1}^{k} \tau_i(N-M_{i-1})^2 (1-e^{-a\tau_i})e^{-a\tau_i} .$$

The iterative procedure for solving these equations requires starting points for both N and c (or a), with the starting point for N being

$$\geq \sum_{i=1}^{k} N_i$$

and the starting point for c or a being >0. A detailed discussion of these and other starting points is given in Section 4.2.2.

The maximum likelihood method finds those values $\hat{N}$ and $\hat{c}$ (or $\hat{a}$) which maximize the logarithm of the likelihood function,

$$\ln L = \ln \prod_{i=1}^{k} \binom{N-M_{i-1}}{N_i} p_i^{N_i} (1-p_i)^{N-M_i} .$$

Due to the computational difficulties and computer-time requirements MLEs were not calculated for the binomial models.

4.1.5 Estimation for the IBM Growth Model

The computerized methods of obtaining estimators for the model of the non-homogeneous Poisson Process described in Section 3.2 follow. Here the number of errors occurring through time t_i, $N(t_i)$, is considered to have a Poisson distribution with mean $m(t_i) = a(1-e^{-bt_i})$ with a and b positive. Also $N(t_0) = 0$ and $t_0 = 0$.

Here the LSEs are those values $\tilde{a}$ and $\tilde{b}$ which minimize the sum

$$S = \sum_{i=1}^{k} \left\{ \left[N(t_i) - N(t_{i-1}) \right] - \left[m(t_i) - m(t_{i-1}) \right] \right\}^2 .$$

In particular, the LSEs are the solutions to the equations:

$$\frac{\partial S}{\partial a} = \frac{\partial S}{\partial b} = 0$$

where

$$\frac{\partial S}{\partial a} = -2 \sum_{i=1}^{k} \left[\left(N(t_i) - N(t_{i-1}) \right) - a \left(e^{-bt_{i-1}} - e^{-bt_i} \right) \right] \left(e^{-bt_{i-1}} - e^{-bt_i} \right)$$

$$\frac{\partial S}{\partial b} = -2 \sum_{i=1}^{k} \left[(N(t_i) - N(t_{i-1})) - a \left(e^{-bt_{i-1}} - e^{-bt_i} \right) \right] a \left(t_i e^{-bt_i} - t_{i-1} e^{-bt_{i-1}} \right)$$

The parameter a was eliminated from the two equations above leaving one equation to solve for $\tilde{b}$ using the Newton-Raphson method. The estimate $\tilde{a}$ is then calculated from the equation:

$$\tilde{a} = \sum_{i=1}^{k} (N(t_i) - N(t_{i-1})) \left(e^{-\tilde{b}t_{i-1}} - e^{-\tilde{b}t_i} \right) \Big/ \sum_{i=1}^{k} \left(e^{-bt_{i-1}} - e^{-\tilde{b}t_i} \right)^2 .$$

The MLEs are those values $\hat{a}$ and $\hat{b}$ which maximize the logarithm of the likelihood function,

$$\ln L = \ln \prod_{i=1}^{k} \left[\frac{\exp\left(- (m(t_i) - m(t_{i-1})) \right) \{ m(t_i) - m(t_{i-1}) \}^{Z_i}}{Z_i \,!} \right]$$

where

$$Z_i = N(t_i) - N(t_{i-1}) , \; i=1, \ldots, k, \; t_0 \equiv 0 .$$

Here the MLEs are the solutions to the equations:

$$\frac{\partial \ell n\, L}{\partial a} = \frac{\partial \ell n\, L}{\partial b} = 0$$

where

$$\frac{\partial \ell n\, L}{\partial a} = \sum_{i=1}^{k} \frac{Z_i}{a} - (1-e^{-bt_k})$$

$$\frac{\partial \ell n\, L}{\partial b} = -a\, t_k\, e^{-bt_k} + \sum_{i=1}^{k} \frac{Z_i\left(t_i\, e^{-bt_i} - t_{i-1}\, e^{-bt_{i-1}}\right)}{\left(e^{-bt_{i-1}} - e^{-bt_i}\right)}$$

The parameter a was eliminated from the two equations above leaving one equation to solve for $\hat{b}$ using the Newton-Raphson method. The estimate $\hat{a}$ was then computed from the equation:

$$\hat{a} = N(t_k)/\left(1-e^{-\hat{b}t_k}\right).$$

4.2 PROPERTIES OF ESTIMATES

4.2.1 Ease of Calculation

Throughout the applications of the models, there were varying degrees of ease of calculation. Often, calculations were simplified by approximating derivatives in order to apply the Newton-Raphson method. This made computer computations easier (as well as less time consuming and expensive), but also gave slightly less accurate estimates. Derivative approximations were used to find estimates in the three parameter Poisson model, both binomial models and the IBM growth model.

The greatest difficulty in implementing a method to find parameter estimates occurred in applying the ML method to the binomial models of Section 3.3. Here finding estimates for the parameter c (or a) which maximized the likelihood function for corresponding integer N's (over all values of

$$N \geq \sum_{i=1}^{k} N_i)$$

proved much too computer time consuming and costly for practical use. This also suggests that testing only integer estimates of N would prove to be impractical for the other models also.

Another computational problem arises in applying a second derivative test to insure attaining the proper extremum for the three parameter Poisson model. These computations become extremely complicated, and may produce inaccurate results when applied to the approximations of the estimators $\hat{N}$, $\hat{\alpha}$, and $\hat{\phi}$ (or $\tilde{N}$, $\tilde{\alpha}$ and $\tilde{\phi}$). This problem may also arise in applying a second derivative test for extrema to two parameter cases.

Specifically, a discrepancy arose in applying a second derivative test to a fit of the Jelinski-Moranda model (ML method) to data set number 6-fd. Here the estimates $\hat{N} = 113,594$ and $\hat{\phi} = 8.95 \times 10^{-6}$ were evaluated to be maximizing estimates of the parameters; while the values $\hat{N} = 113,594$ and $\hat{\phi} = 9.0 \times 10^{-6}$ were considered to act as a saddle point. This difference in round-off error on the order of 10^{-8} may illustrate the existence of severe precision problems in computing parameter estimates.

4.2.2 Convergence Difficulties

Throughout the analysis difficulties in achieving convergence of the estimates were encountered. One of the most acute problems is choosing satisfactory starting points as inputs (required to initiate the Newton-Raphson method of solution). A basic problem with starting points is that none of the literature discussing the various software reliability models contain any conventions for obtaining starting points, although iterative processes are usually cited for use in solving the relevant equations.

The most severe cases of starting point problems occurs in the models requiring two simultaneously input starting points. Some cases with this problem are shown in Table 4.2.2.1.

This table shows that small differences in the input starting points may result in vast differences of attaining convergence. This is especially true for the values in the last two rows of Table 4.2.2.1.

Observation shows that the choice of the starting point is more critical in applying the ML method to a particular model than in applying the LS method. This may be due to the need to solve less complicated equations in the latter case.

The possibility of the existence of more than one set of solutions to the likelihood (or sum of squares) equations is seen most clearly in applications of the IBM model. Here it was found that both the LS and ML method could lead to convergence to two (or more) different sets of estimates – one set having a finite positive value for $\tilde{a}$ (or $\hat{a}$), and a positive value for $\tilde{b}$ ($\hat{b}$); the other having the iterative approximations for $\tilde{a}$ ($\hat{a}$) increasing without bound, the values for $\tilde{b}$ ($\hat{b}$) tending to zero, and the product $\tilde{a}\tilde{b}$ ($\hat{a}\hat{b}$) approaching a constant value. Upon encountering the latter case, changing starting points could sometimes result in the former case of finite, non-zero estimates of the parameters. This use of change in starting points is illustrated in Table 4.2.2.2.

The choice of starting points appeared to be the least critical in applying the LS method to the Jelinski-Moranda and Shick-Wolverton models of Section 3.1. Because of this situation, in trying to fit a particular data set to the models, it may be helpful to try finding parameter estimates for these models before applying the other models

Table 4.2.2.1. Starting Point Analysis for the Binomial Model

BINOMIAL I

Starting Points		Result
$\tilde{N}$	$\tilde{c}$	
500	15	Non-Convergence
550	25	Non-Convergence
450	200	Non-Convergence
500	15	Convergence to invalid estimates
499	75	CONVERGENCE TO VALID ESTIMATES

BINOMIAL II

Starting Points		Result
$\tilde{N}$	$\tilde{a}$	
3600	0.1	Program interrupt due to error
3500	0.01	Convergence to invalid estimates
3600	0.009	Convergence to invalid estimates
3599	0.008	CONVERGENCE TO VALID ESTIMATES

(or MLE method). In doing so, more efficient starting points for the estimates of the parameter N for the other models may be suggested.

4.2.3 Types of Convergence Observed

Different types of convergence were encountered for the various models. Although monotone convergence (e.g. usually increasing for estimates of N) was most common, the path of the iterations was not always predictable. Oscillation was a common occurrence in applying the ML method of seeking estimates for the IBM model of Section 3.2. Here when estimates of $\hat{b}$ tended toward zero, oscillation with changes in sign about zero often resulted. Notice that this shows there is no practical method of choosing starting points which will always result in positive estimates for $\hat{a}$ and $\hat{b}$.

Table 4.2.2.2. Starting Point Results for the LSEs for the IBM Model – Data Set 14-fd

Starting Point ($\tilde{b}$)	$\tilde{a}$	$\tilde{b}$
0.9	-2.11	-0.6025
0.01	1.52×10^{8}	2.15×10^{-7}
0.55	353.70	0.3510

Results for Data Set 14-fd show different starting points for $\tilde{b}$ to yield widely varied parameter estimates. Using the starting estimate of 0.55 for $\tilde{b}$, however, appears to result in the most appropriate parameter estimates.

The only model found to result in convergence at all times is the IBM model of Section 3.2. Here both the LS and ML methods always gave parameter estimates. The probable explanation for this "guaranteed" convergence is that the iterations of only one parameter estimate (those of $\tilde{b}$ or $\hat{b}$) are compared in order to determine convergence. This results in the case where the estimates of $\tilde{a}$ ($\hat{a}$) diverge (increase in absolute value without bound) being considered "convergent". However, other models treat similar cases (e.g. estimates of $\tilde{N}$ increasing without bound, while estimates of $\tilde{c}$ converge to a finite value) as not being convergent.

Convergence criteria should be standardized for all of the estimates and the models. Without a uniform treatment of concepts such as convergence it can be misleading to compare properties (such as frequency of convergence) of the various models.

Even with convergence achieved, the resulting set of parameter estimates may not actually be the desired extremum. A listing of the actual numbers of proper extrema found for each model is shown in Table 4.3.1.

Second derivative testing did show the existence of some saddle points as estimates for several models. Subsequent computer testing could not always result in convergence to other parameter estimates for these data sets, leaving the question of whether an actual extremum can be found still open.

However, the validity of the second derivative test on the resulting approximations of the parameter estimates should not be considered absolute, as illustrated by the previously stated example (given in Section 4.2.1) where small round-off error for the estimates led to two different conclusions for the test.

4.2.4 Convergence to Invalid Estimates

There is also a problem of tendencies toward convergence to parameter estimates (even proper extrema) which violate the model assumptions.

For most models (all but the IBM model of Section 3.2), convergence to estimates having $\tilde{N}$ ($\hat{N}$) less than the total number of errors observed,

$$\left(\sum_{i=1}^{k} N_i\right)$$

occurred frequently. These estimates, however, were not considered to be fits of the model.

In the three-parameter GPM of Section 3.1, convergence with the resulting estimate $\tilde{\alpha} < 0$ (or $\hat{\alpha} < 0$) was common. This would mean smaller expected error rates over longer time intervals. Nonetheless, these results were considered to be plausible fits. It should be pointed out that changes of some data bases from time units of days to units of hours changed only the estimates of ϕ for fits to the Poisson models of Section 3.1. The remaining estimates of N (and α) were unaltered. That is, they were invariant under changes of scale.

Finally, the oscillation sometimes encountered in applying the ML method to the IBM model of section 3.2 sometimes resulted in negative estimates for both $\hat{a}$ and $\hat{b}$. This contradicts the assumption that

$$a = \lim_{t \to \infty} E(N(t))$$

the expected total number of errors to be found in the software. This problem, however, could usually be eradicated by trying other starting points.

4.3 EXISTENCE, UNIQUENESS AND CONVERGENCE PROPERTIES OF THE ESTIMATORS

The statistical and numerical methods for obtaining parameter estimators described in Section 4.1 may not always yield usable results. First, there may be conditions under which the statistical theory leads to a situation in which the estimator does not exist. Estimators which are not unique are also a problem. That is, the extreme value of the function may result from different sets of parameter estimates. For example, in the case of the MLE's, there may be more than one value of the estimator which yields an identical value for the maximum. In this situation there is no way to know which estimate to choose, or even if all maximizing estimates have been found. Finally, the numerical methods employed may not always lead to convergence.

4.3.1 Existence of the Estimators

If an estimator fails to exist at all times, there will be instances where another estimator (which does exist) must be used instead. The limitations of an estimator which may not exist make it impossible to rely on these estimators for all samples.

In the J-M model described in Section 3.1, for example, the general criteria for existence of the maximum likelihood estimator have not been established. However, a specific case where the MLE of N in the J-M model does not exist has been found.

Consider the case of only two positive time intervals τ_1 and τ_2, with exactly one error in each interval. Let c be the positive constant

$$c = \tau_2 / \tau_1$$

With the parameter ϕ eliminated, the likelihood function of N for this case becomes

$$L(N) = \frac{4c\,N(N-1)}{[N + c(N-1)]^2 e^2}$$

where e is the usual base of natural logarithms.

The ratio, c, of the sizes of the two time intervals is crucial here. Notice that any estimator of N must be ≥ 2.

i) If $0 < c \leq 1$, L(N) (as shown in Figure 4.3.1a) is a monotone increasing function, and therefore the MLE does not exist. In this example, then, if the error-free time to the second observation is no larger than that to the first, the maximum likelihood method of finding an estimator for N will not yield a solution.

ii) Further calculations show that if $c \geq \sqrt{3}$, the integer value (since N is an integer, so must its MLE, $\hat{N}$, be) which maximizes the likelihood function is $\hat{N} = 2$, the total number of errors already observed (see Figure 4.3.1b). This is not an invalid estimate, but may be somewhat impractical. Here, with only two errors observed it is, in most cases, unlikely that they represent all of the errors.

iii) Finally, if $1 < c < \sqrt{3}$ the MLE will be an integer greater than 2, depending on the exact value of c. That is, the MLE exists and indicates that undetected errors still remain (see Figure 4.3.1c).

The establishment of existence criteria for the MLE in this example was due largely to the simplicity of the data base (i.e. two intervals). In more complicated cases, existence of the estimators has yet to be established. However, the above result must not be underrated. Although obtained under conditions of a very small data base, it establishes that there is no guarantee of the existence of the MLE in any situation.

Valid estimators also will not exist if their resulting values are contrary to the model assumptions (these assumptions are discussed in Section 3). Many such contradictions were encountered in actually testing the data bases of Section 2.

One contradiction occurs when the estimator of the total number of errors in the software, $\hat{N}$ (or $\tilde{N}$), is smaller than the actual number of errors already observed,

$$\sum_{i=1}^{k} N_i .$$

Both the ML and LS methods were subject to this problem. More severely, even negative estimates for model parameters assumed to be positive were obtained. This is not an uncommon situation in statistics. The moment estimators often result in negative estimates for the variance. This is one reason why moment estimators are rarely used.

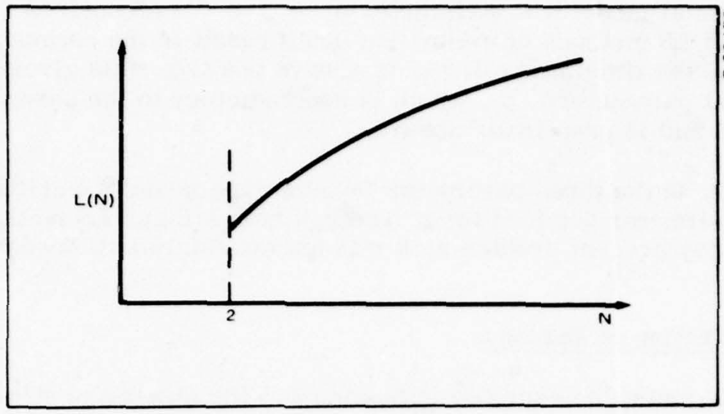

Figure 4.3.1a. Graph of L(N) for $0 < c \leq 1$

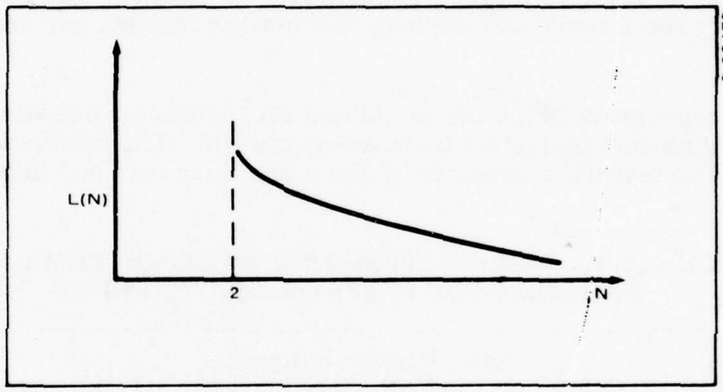

Figure 4.3.1b. Graph of L(N) for $c \geq \sqrt{3}$

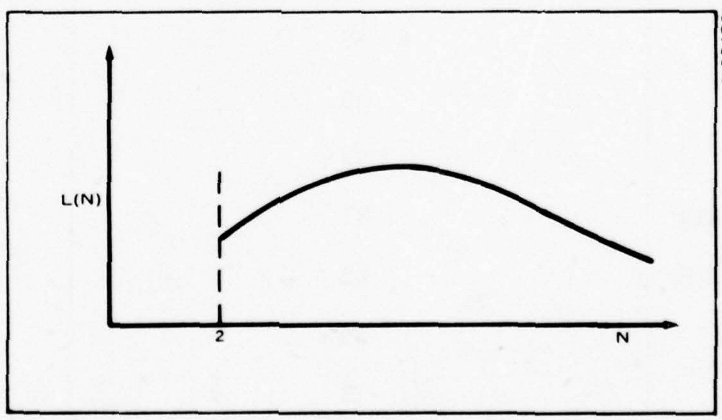

Figure 4.3.1c. Graph of L(N) for $1 < c < \sqrt{3}$

In the binomial model with parameter $p_i = 1-e^{-a\tau_i}$ (described in Section 3.3), both the ML and LS methods of estimation could result in the estimator $\hat{a}$ (or $\tilde{a}$) being negative. Since the time interval, τ_i, is always positive, this gives a negative value for the binomial parameters, p_i, which is contradictory to the assumptions of the binomial model and is practically absurd.

Frequently, in the three parameter Poisson type model described in Section 3.1, the resulting estimator obtained for α (through both ML and LS methods) was negative. Although this may present problems, it was not considered an invalid result.

4.3.2 Determination of Extrema

Most of the methods described in Section 4.1 for obtaining estimators used only a first derivative test, which does not necessarily yield the proper extremum. That is, a maximum value rather than a minimum value may be found in using this method of finding the LSEs (similarly, a minimum found as the MLEs). Saddle points (critical values satisfying the first derivative test, but neither minima nor maxima) could also occur.

A computer program which further tested all resulting estimators for the proper extremum using second derivative tests was designed. The results are shown in Table 4.3.1. The test did show some of the estimators obtained to be saddle points.

TABLE 4.3.1. VERIFICATION OF PROPER EXTREMA IN LSEs AND MLEs FOR CASES OF CONVERGENCE

Model	No. Proper Extrema (Minimum for LSE, Maximum for MLE)	No. Saddle Points
IBM (LSE)	50	4
IBM (MLE)	46	7
J-M (LSE)	38	0
J-M (MLE)	39	3
Binomial I (LSE)	37	0
Binomial II (LSE)	32	0
S-W (LSE)	20	2
S-W (MLE)	23	0
GPM (LSE)	36	0
GPM (MLE)	38	2

It is possible that, when a saddle point or improper extremum results, a change in starting points could yield the proper extremum. This situation is discussed further in Section 4.2.2.

4.3.3 Convergence of the Newton-Raphson Method

In the actual use of the Newton-Raphson method, convergence of the estimators to finite values could not always be obtained. The major problem seemed to be in finding successful starting points for the parameter estimates as inputs to the programs. In general, no real guidelines were found.

Some starting points led to divergence of the estimates of $\hat{N}$ (or $\tilde{N}$), i.e. the successive approximations of $\hat{N}$ (or $\tilde{N}$) would increase steadily without bound. This could sometimes be corrected by decreasing the initial value of N.

Some starting points would give convergence to invalid values for the estimators, while changing the starting points sometimes gave valid estimates of the parameters. (See Tables 4.2.2.1, 4.2.2.2). Often in the Poisson models of Section 3.1 a data set resulting in a negative value for $\hat{N}$ (or $\tilde{N}$) could yield a valid estimate for N if its starting point was sufficiently increased. The possibility exists that different starting points can lead to convergence to more than one valid set of estimators for a single set of data. In such cases there was no measure found guaranteeing which values were more appropriate.

Particular difficulty in finding useful starting points for the parameter estimates occurred in the three parameter Poisson model of Section 3.1 and the least squares version of the binomial models of Section 3.3. Here both starting points (initial values for the estimates of N, α; N, c; or N, a) must be coordinated in order to succeed in obtaining convergence for the estimates. In the binomial case, even if convergence has been obtained, reusing those estimators as input starting points for the same model and method on the same data set will not always result in convergence again. This definitely presents problems in choosing initial values to input.

For the J-M and S-W models of Section 3.1, convergence seems somewhat easier to obtain when using the least squares method than with the maximum likelihood method. Convergence in those computer programs using the LS method was usually faster, i.e. fewer iterations were needed in the convergent sequence of parameter estimates.

Convergence also is frequently more direct in the programs using the least squares method, e.g. a monotone convergent sequence of parameter estimates rather than an oscillating sequence.

4.3.4 Uniqueness of the Estimators

The uniqueness of either the maximum likelihood or least squares parameter estimators cannot be insured as the following example illustrates.

Consider the MLE of N for the particular case of the J-M model of Section 3.1. Here, assume two positive time intervals τ_1 and τ_2 with the positive constant

$$c = \tau_2/\tau_1$$

as before. Assume no errors observed in the first time interval, τ_1, and one error detected in the second interval, τ_2. With the parameter ϕ eliminated, the likelihood function evaluated for this case becomes

$$L(N) = \frac{c}{(c+1)\,e}$$

which is a constant independent of N. Therefore, since <u>any</u> value of N maximizes the likelihood function, the MLE of N in this case is not unique whatever be the ratio c.

The possibility of a multimodal likelihood function (or more than one local minimum of the sum of squares function S) cannot be eliminated, either. The present numerical techniques of finding the parameter estimators cannot accommodate this situation in finding an overall maximum (or minimum) value. Specifically, in the present ML method of finding parameter estimators for the binomial models (maximizing the likelihood function for each integer value of N) multimodal situations were observed. Therefore, the required size of a sufficiently large testing range for N could not be established.

No method of guaranteeing the uniqueness of either the MLEs or LSEs has been found. The numerical techniques presently used in finding these estimators can detect only one set of parameter estimates which satisfy the relevant equations, and do not eliminate the existence of other estimates which may yield the same (or possibly better) value of the likelihood function, L, or sum of squares, S. Without a guarantee of uniqueness, it may be advisable to continue testing a model for a single data set, even after a suitable set of parameter estimates has been obtained.

If criteria for the existence of the estimators can be established, some of the difficulties in obtaining covergence for the estimators might be eliminated. One idea is that the observed rate of increase (or decrease) of errors could possibly be used to establish bounds for values of the estimators. This could aid in the selection of the proper estimate in a multinodal (more than one maximum) situation or conserve efforts in cases where nonexistence of the estimators has been determined. This idea has not been studied here. Additionally, such criteria could possibly aid in finding suitable starting points for the estimates. In particular, determining bounds on the estimators could lead to a coordination of the input values necessary in the models requiring two starting points.

It is significant to notice that, at present, when parameter estimators cannot be obtained for a particular data set, there is no way to distinguish between the nonexistence of these estimators and failures in the numerical method of solution (e.g., improper starting points).

4.4 SMALL AND LARGE SAMPLE PROPERTIES OF THE ESTIMATORS

4.4.1 Lar Sample Results: MLEs

4.4.1.1 The GPM, S-W Model, and J-M Model

True asymptotic results for the MLEs and LSEs for the GPM, J-M and S-W models are not valid unless the assumption is made that not all of the N initial errors can ever be removed. That is, we must assume that there are numbers $\underline{M}$ and $\overline{M}$ such that for all i, $0 \le \underline{M} \le M_{i-1} \le \overline{M} < N$. If this is not the case, the successive observed errors $N_1, N_2, \ldots$ will eventually degenerate to the value zero. Indeed, no matter how much debugging time is spent, errors still occur although they are usually rare in a delivered software package. The existence of the number $\overline{M}$ then appears reasonable and renders proper asymptotic analysis of the software models mentioned above possible.

As a further assumption it is required that there are times $\underline{t}$ and $\overline{t}$ such that $0 < \underline{t} \le \tau_i \le \overline{t} < \infty$ for all i.

As mentioned in Section 3.1 the likelihood function for the sample $N_1, \ldots, N_k$ is

$$\ell(N_1, \ldots, N_k;\ \phi, N, \alpha) = \prod_{i=1}^{k} \frac{\left(\phi(N - M_{i-1})\tau_i^{\alpha}\right)^{N_i}}{N_i!} e^{-\phi(N-M_{i-1})\tau_i^{\alpha}}$$

for the GPM ($\alpha = 1$ corresponds to J-M, $\alpha = 2$ corresponds to S-W). Denoting the natural log of the likelihood function by $K \equiv K(N_1, \ldots, N_k;\ \phi, N, \alpha)$ it is seen from Section 4.1.3 that

$$\frac{\partial K}{\partial N} = \sum_{i=1}^{k} \frac{N_i}{(N-M_{i-1})} - \phi \sum_{i=1}^{k} \tau_i^{\alpha}$$

$$\frac{\partial K}{\partial \phi} = \sum_{i=1}^{k} \frac{N_i}{\phi} - \sum_{i=1}^{k} (N-M_{i-1})\tau_i^{\alpha}$$

$$\frac{\partial K}{\partial \alpha} = \sum_{i=1}^{k} N_i \ln \tau_i - \sum_{i=1}^{k} \phi(N-M_{i-1})\tau_i^{\alpha} \ln \tau_i .$$

The classical theory of maximum likelihood (cf. (I,3) for example) states that under certain conditions, the distribution of $(\hat{N}, \hat{\phi}, \hat{\alpha})^T$ (T denotes transpose) is asymptotically multivariate normal with mean $(N, \phi, \alpha)^T$ and covariance matrix

$$C^{-1} = \begin{pmatrix} -E\left(\frac{\partial^2 K}{\partial N^2}\right) & -E\left(\frac{\partial^2 K}{\partial N \partial \phi}\right) & -E\left(\frac{\partial^2 K}{\partial N \partial \alpha}\right) \\ -E\left(\frac{\partial^2 K}{\partial N \partial \phi}\right) & -E\left(\frac{\partial^2 K}{\partial \phi^2}\right) & -E\left(\frac{\partial^2 K}{\partial \phi \partial \alpha}\right) \\ -E\left(\frac{\partial^2 K}{\partial N \partial \alpha}\right) & -E\left(\frac{\partial^2 K}{\partial \phi \partial \alpha}\right) & -E\left(\frac{\partial^2 K}{\partial \alpha^2}\right) \end{pmatrix}^{-1}$$

where the partial derivatives are evaluated at the actual values of N, ϕ, and α. This result is true for the GPM and can be proven in a straight forward manner (see sections dealing with the derivation of asymptotic properties of LSEs for the general techniques involved). Performing the partial differentiations and taking expections gives

$$C^{-1} = \begin{pmatrix} \sum\limits_{i=1}^{k} \frac{\phi \tau_i^{\alpha}}{(N-M_{i-1})} & \sum\limits_{i=1}^{k} \tau_i^{\alpha} & \phi \sum\limits_{i=1}^{k} \tau_i^{\alpha} \ell n \tau_i \\ & \sum\limits_{i=1}^{k} \frac{(N-M_{i-1}) \tau_i^{\alpha}}{\phi} & \sum\limits_{i=1}^{k} (N-M_{i-1}) \tau_i^{\alpha} \ell n \tau_i \\ & & \sum\limits_{i=1}^{k} (N-M_{i-1}) \tau_i^{\alpha} (\ell n \tau_i)^2 \end{pmatrix}^{-1}$$

where the lower triangular portion is left blank since the entries are the same as their diagonal opposites due to symmetry. Since N, ϕ, and α are unknown C^{-1} may be estimated by replacing N, ϕ, and α by $\hat{N}$, $\hat{\phi}$, and $\hat{\alpha}$. It should be noted, though, that doing this makes C^{-1} a random variable and it is no longer correct to say that

$$\begin{pmatrix} \hat{N} - N \\ \hat{\phi} - \phi \\ \hat{\alpha} - \alpha \end{pmatrix} \text{ is approximately normally}$$

distributed with mean 0 and covariance matrix $\hat{C}^{-1}$ for large k, where $\hat{C}^{-1}$ is C^{-1} with N, ϕ and α replaced by $\hat{N}$, $\hat{\phi}$ and $\hat{\alpha}$. All one can say is that $\hat{C}^{-1}$ is the maximum likelihood estimator of C^{-1}.

The matrix C^{-1} is difficult to compute algebraically and it is recommended here that estimators $\hat{C}^{-1}$ be computed via a digital computer. The estimators of the variances and covariances of $\hat{N}$, $\hat{\phi}$ and $\hat{\alpha}$ would then be

$$\mathrm{Var}(\hat{N}) \overset{a}{=} \hat{C}^{-1}_{11}$$

$$\mathrm{Var}(\hat{\phi}) \overset{a}{=} \hat{C}^{-1}_{22}$$

$$\mathrm{Var}(\hat{\alpha}) \overset{a}{=} \hat{C}^{-1}_{33}$$

$$\mathrm{Cov}(\hat{N}, \hat{\phi}) \overset{a}{=} \hat{C}^{-1}_{12} = \hat{C}^{-1}_{21}$$

$$\mathrm{Cov}(\hat{N}, \hat{\alpha}) \overset{a}{=} \hat{C}^{-1}_{13} = \hat{C}^{-1}_{31}$$

$$\mathrm{Cov}(\hat{\phi}, \hat{\alpha}) \overset{a}{=} \hat{C}^{-1}_{23} = \hat{C}^{-1}_{32}$$

where

$$\hat{C}^{-1}_{ij}$$

denotes the entry in the i^{th} row and j^{th} column of $\hat{C}^{-1}$ and where "$\overset{a}{=}$" means asymptotically equal, in probability, as $k \to \infty$.

For the S-W and J-M models, the results are the same except α is no longer estimated and the last row and last column of C^{-1} are eliminated. In these cases, the inversion is easily carried out and it follows that

$$\mathrm{Var}(\hat{N}) \overset{a}{=} \sum_{i=1}^{k} \frac{(N-M_{i-1})\,\tau_i^{\alpha}}{\phi} \Big/ D$$

$$\mathrm{Var}(\hat{\phi}) \overset{a}{=} \sum_{i=1}^{k} \frac{\phi\,\tau_i^{\alpha}}{(N-M_{i-1})} \Big/ D$$

$$\mathrm{Cov}(\hat{N}, \hat{\phi}) \overset{a}{=} -\sum_{i=1}^{k} \tau_i^{\alpha} \Big/ D$$

where

$$D = \sum_{i=1}^{k} (N-M_{i-1}) \tau_i^{\alpha} \sum_{i=1}^{k} \frac{\tau_i^{\alpha}}{(N-M_{i-1})} - \left(\sum_{i=1}^{k} \tau_i^{\alpha} \right)^2 ,$$

and $\alpha = 1$ for J-M, $\alpha = 2$ for S-W. As before, since N and ϕ are unknown, maximum likelihood estimators of these quantities are obtained by replacing N and ϕ by $\hat{N}$ and $\hat{\phi}$. To verify these results for the J-M and S-W models, the sampling distributions of N, along with their asymptotic variances and covariances, were simulated based on true values of $N = 300$ and $\phi = 0.01$. This was accomplished by taking $k = 25$ intervals, $\tau_i = 10$ for all i, and $M_{i-1} = 10\ (i - 1)$, $i = 1, 2, \ldots, k$. Five hundred sets of observations $N_1, N_2, \ldots, N_{25}$ were generated using a Poisson random number generator. For each of the 500 sets, $\hat{N}$, $\hat{\phi}$ were computed and stored. Histograms of the 500 $\hat{N}$'s and $\hat{\phi}$'s were printed along with a tabulation of the class intervals. Following the class interval tabulations are the sample standard deviation and mean (of the respective parameter estimator). Finally, the sample covariance between the 500 $\hat{N}$'s and $\hat{\phi}$'s is printed along with the asymptotic variances and covariances derived in this section computed for $N = 300$, $M_{i-1} = 10\ (i - 1)$, $\phi = 0.01$, $\tau_i = 10$ for all i, and $k = 25$. These simulations and calculations are presented in Figure 4.4.1.1 for the J-M and S-W models. As shown, even for such small k the asymptotic results are quite accurate and approximate normality is valid.

4.4.1.2 IBM Model

In the proof of asymptotic normality of MLEs the key observation is that the vector of partial derivatives of the log of the likelihood function with respect to the parameters estimated is asymptotically normally distributed. In the case of the IBM model (letting K denote the log of the likelihood function) it is seen from Section 4.1.5 that

$$\frac{\partial K}{\partial a} = \sum_{i=1}^{k} Z_i/a - \left(1 - e^{-bt_k}\right)$$

$$= \frac{N(t_k)}{a} - \left(1 - e^{-bt_k}\right)$$

(Since $N(t_0) \equiv 0$ and $Z_i = N(t_i) - N(t_{i-1})$) and

$$\frac{\partial K}{\partial b} = \sum_{i=1}^{k} \frac{Z_i \left(t_i e^{-bt_i} - t_{i-1} e^{-bt_{i-1}}\right)}{\left(e^{-bt_{i-1}} - e^{-bt_i}\right)} - a\, t_k e^{-bt_k}$$

82755-5

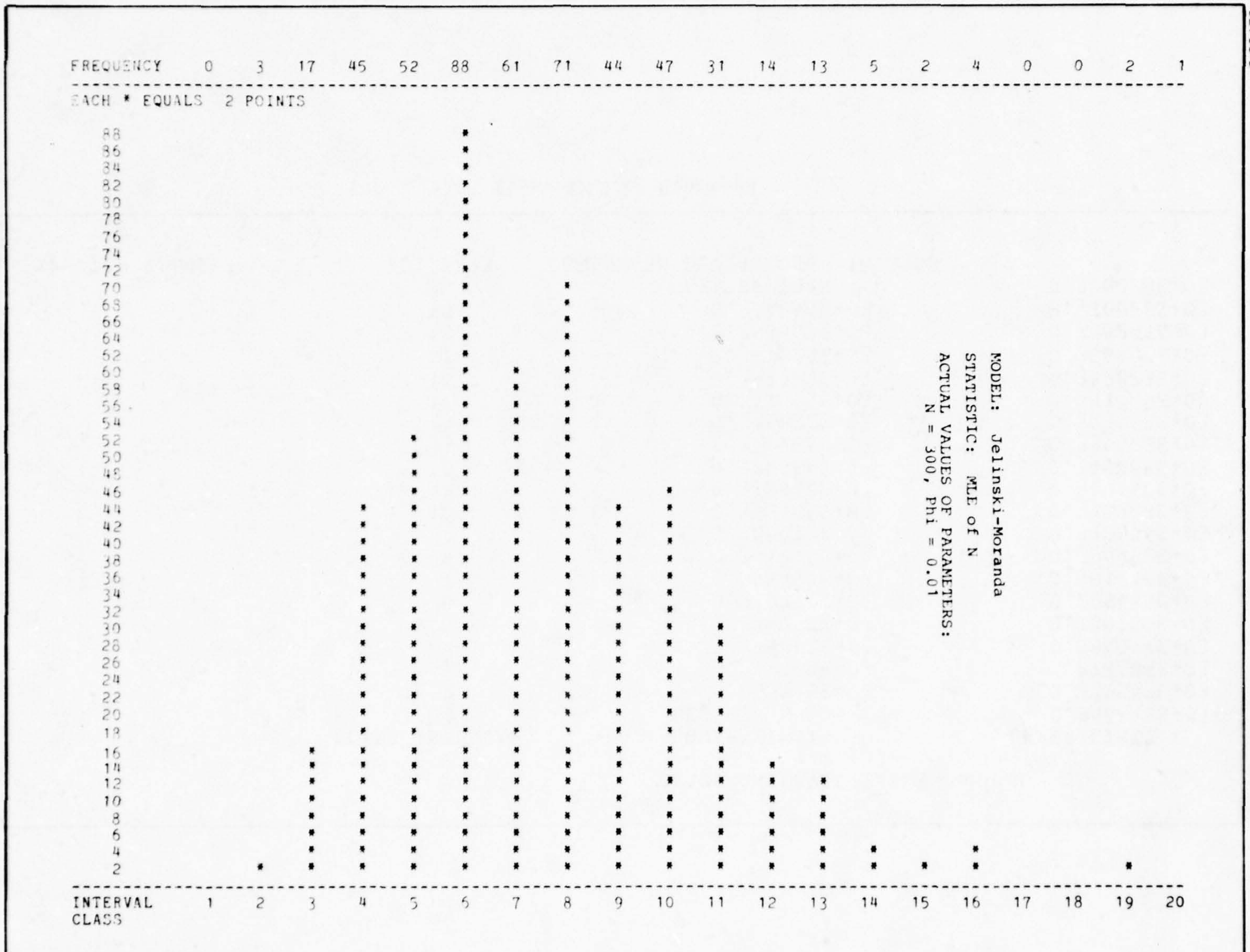

Figure 4.4.1.1. Simulation of the Distributions of the MLEs for the J-M and S-W Models

82755-6

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.266815E+03
2	0.266815E+03	0.272549E+03
3	0.272549E+03	0.278283E+03
4	0.278283E+03	0.284017E+03
5	0.284017E+03	0.289750E+03
6	0.289750E+03	0.295484E+03
7	0.295484E+03	0.301218E+03
8	0.301218E+03	0.306952E+03
9	0.306952E+03	0.312686E+03
10	0.312686E+03	0.318420E+03
11	0.318420E+03	0.324154E+03
12	0.324154E+03	0.329888E+03
13	0.329888E+03	0.335622E+03
14	0.335622E+03	0.341355E+03
15	0.341355E+03	0.347089E+03
16	0.347089E+03	0.352823E+03
17	0.352823E+03	0.358557E+03
18	0.358557E+03	0.364291E+03
19	0.364291E+03	0.370028E+03
20	GREATER THAN	0.370028E+03

EXPECTED VALUE= 301.6423 STANDARD DEVIATION= 16.2986

Figure 4.4.1.1. Continued

82755-7

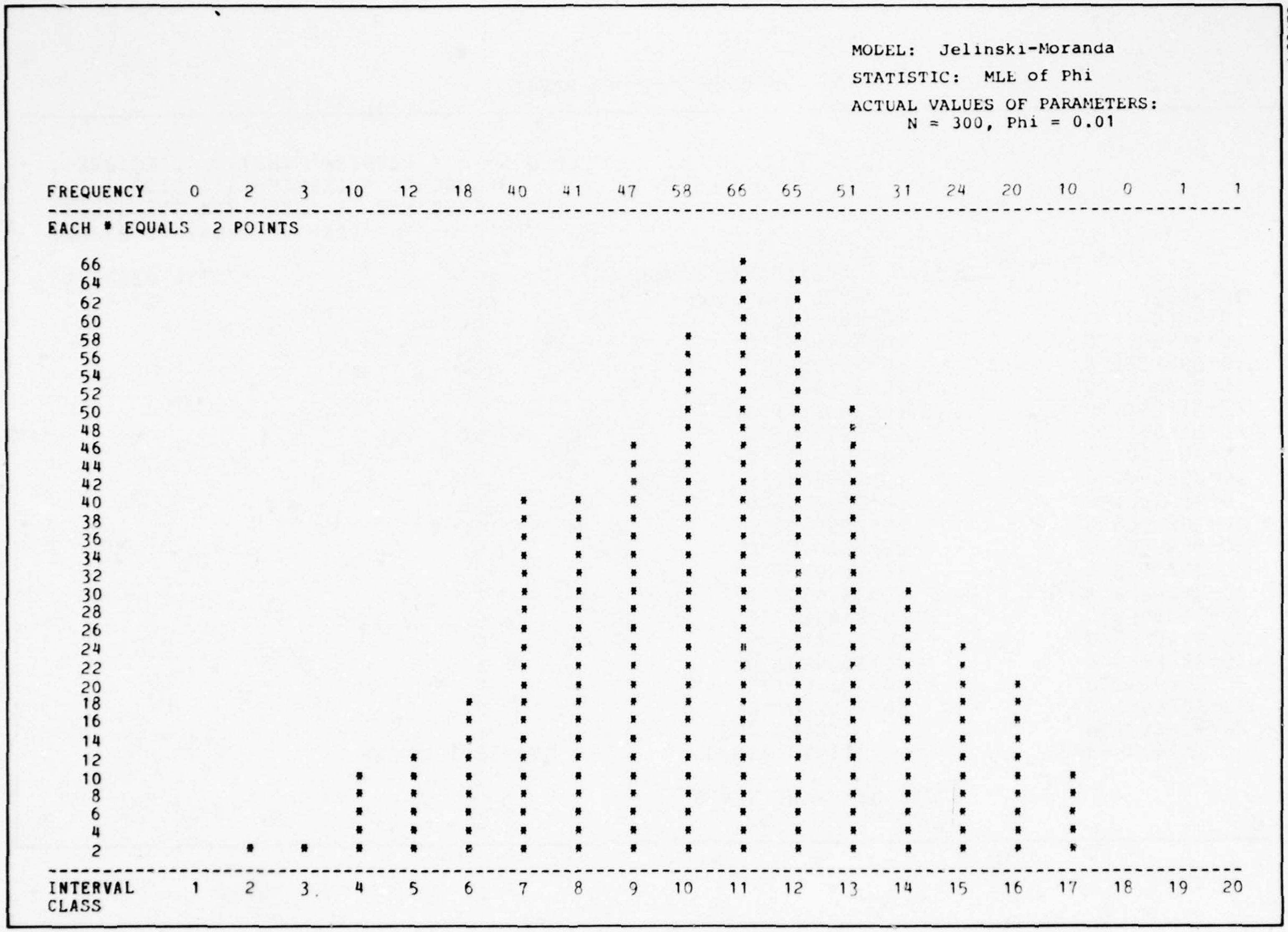

Figure 4.4.1.1. Continued

82755-8

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.703743E-02
2	0.703743E-02	0.736220E-02
3	0.736220E-02	0.768696E-02
4	0.768696E-02	0.801172E-02
5	0.801172E-02	0.833649E-02
6	0.833649E-02	0.866125E-02
7	0.866125E-02	0.898601E-02
8	0.898601E-02	0.931078E-02
9	0.931078E-02	0.963554E-02
10	0.963554E-02	0.996030E-02
11	0.996030E-02	0.102851E-01
12	0.102851E-01	0.106098E-01
13	0.106098E-01	0.109346E-01
14	0.109346E-01	0.112594E-01
15	0.112594E-01	0.115841E-01
16	0.115841E-01	0.119089E-01
17	0.119089E-01	0.122336E-01
18	0.122336E-01	0.125584E-01
19	0.125584E-01	0.128832E-01
20	GREATER THAN	0.128832E-01

EXPECTED VALUE= 0.0100 STANDARD DEVIATION= 0.0010

SAMPLE COV(NHAT,PHIHAT) = -0.0146
ASYMPTOTIC VAR(NHAT) = 243.8889
ASYMPTOTIC VAR(PHIHAT) = 0.00000107
ASYMPTOTIC COV(NHAT,PHIHAT) = -0.0143

Figure 4.4.1.1. Continued

82755-9

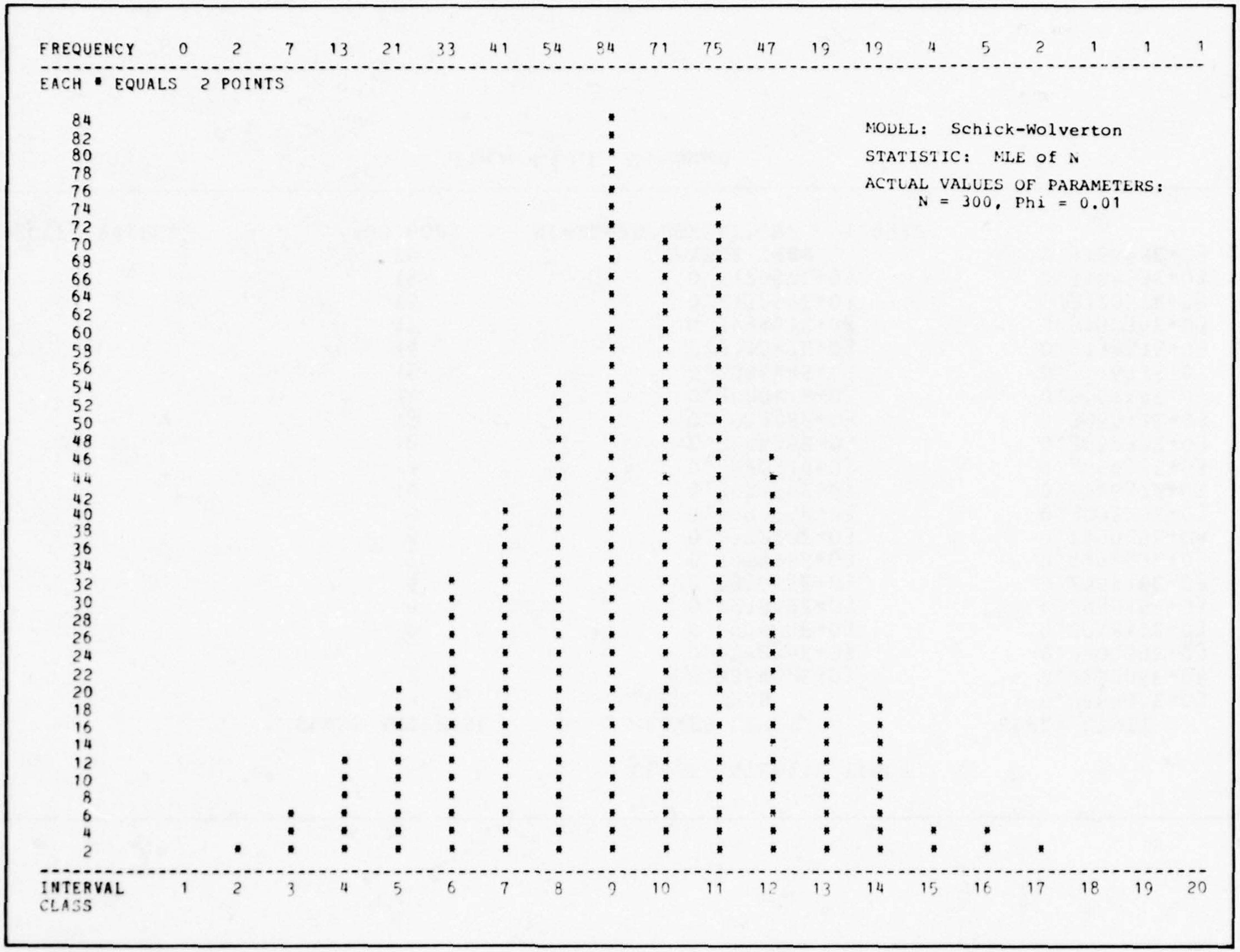

Figure 4.4.1.1. Continued

82755-10

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.286402E+03
2	0.286402E+03	0.288206E+03
3	0.288206E+03	0.290009E+03
4	0.290009E+03	0.291812E+03
5	0.291812E+03	0.293615E+03
6	0.293615E+03	0.295418E+03
7	0.295418E+03	0.297222E+03
8	0.297222E+03	0.299025E+03
9	0.299025E+03	0.300828E+03
10	0.300828E+03	0.302631E+03
11	0.302631E+03	0.304435E+03
12	0.304435E+03	0.306238E+03
13	0.306238E+03	0.308041E+03
14	0.308041E+03	0.309844E+03
15	0.309844E+03	0.311647E+03
16	0.311647E+03	0.313451E+03
17	0.313451E+03	0.315254E+03
18	0.315254E+03	0.317057E+03
19	0.317057E+03	0.318863E+03
20	GREATER THAN	0.318863E+03

EXPECTED VALUE= 300.6089 STANDARD DEVIATION= 4.9932

Figure 4.4.1.1. Continued

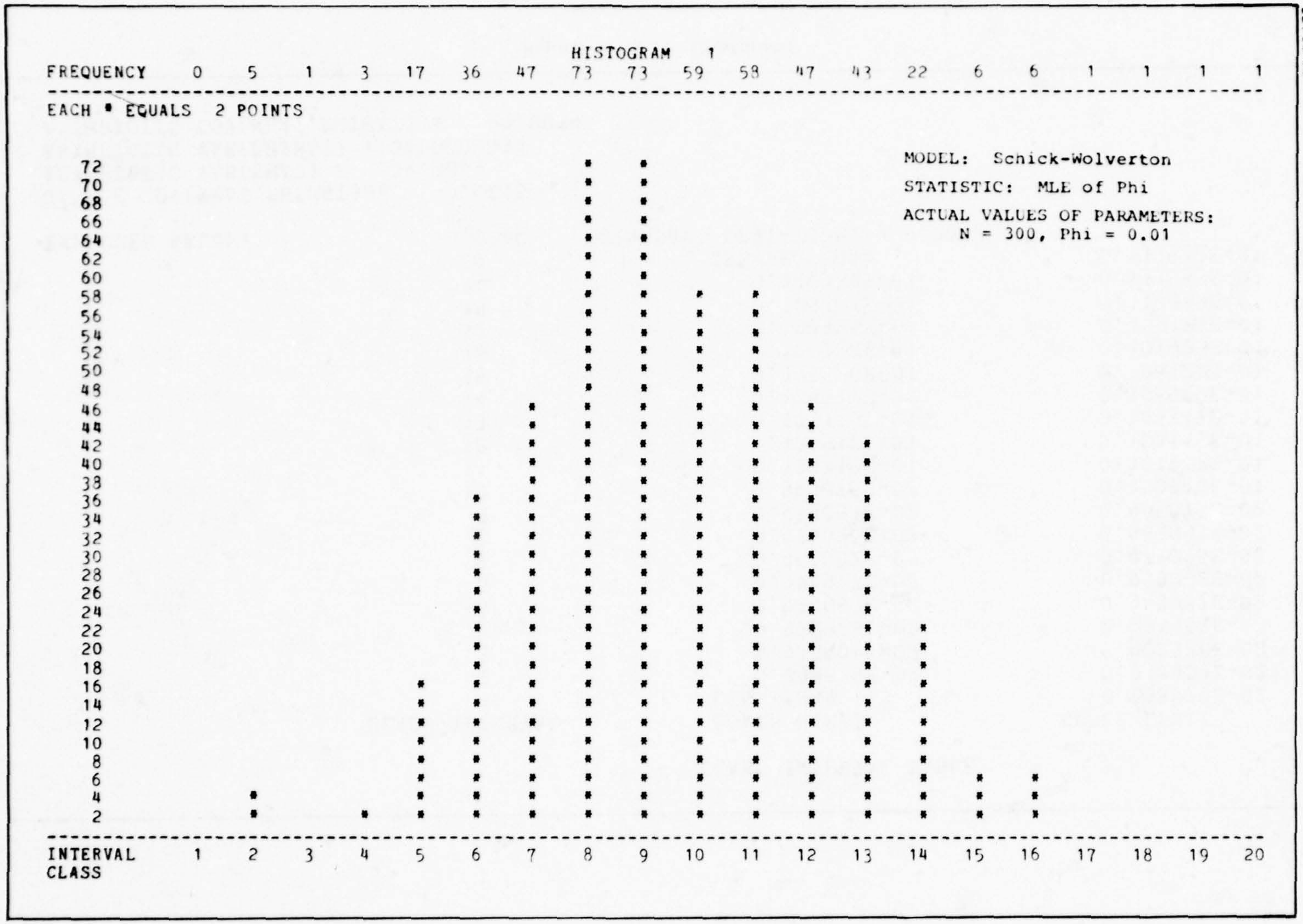

Figure 4.4.1.1. Continued

82755-12

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.898879E-02
2	0.898879E-02	0.910904E-02
3	0.910904E-02	0.922928E-02
4	0.922928E-02	0.934953E-02
5	0.934953E-02	0.946977E-02
6	0.946977E-02	0.959002E-02
7	0.959002E-02	0.971026E-02
8	0.971026E-02	0.983051E-02
9	0.983051E-02	0.995075E-02
10	0.995075E-02	0.100710E-01
11	0.100710E-01	0.101912E-01
12	0.101912E-01	0.103115E-01
13	0.103115E-01	0.104317E-01
14	0.104317E-01	0.105520E-01
15	0.105520E-01	0.106722E-01
16	0.106722E-01	0.107925E-01
17	0.107925E-01	0.109127E-01
18	0.109127E-01	0.110330E-01
19	0.110330E-01	0.111533E-01
20	GREATER THAN	0.111533E-01

EXPECTED VALUE= 0.0100 STANDARD DEVIATION= 0.0003

SAMPLE COV(NHAT,PHIHAT) = -0.0015
ASYMPTOTIC VAR(NHAT) = 24.3889
ASYMPTOTIC VAR(PHIHAT) = 0.00000011
ASYMPTOTIC COV(NHAT,PHIHAT) = -0.0014

Figure 4.4.1.1. Continued

If we assume $0 \equiv t_0 < t_1 < \cdots < t_{k+1} \ldots.$ then it is obvious that with probability 1,

$$\frac{\partial K}{\partial a} \to \frac{N(\infty)}{a} - 1 \quad \text{as} \quad k \to \infty$$

(see Section 3.2 for a discussion of $N(\infty)$).

Now $N(\infty)$ has a Poisson distribution with mean a, and hence $\partial K/\partial a$ is not normally distributed asympototically so that the classical theory for asymptotic properties of maximum likelihood estimators cannot hold exactly. However, for large a (a is usually quite large since it is the expected total number of errors observable in infinite time) $N(\infty)$ is approximately normally distributed with mean a and variance a. So, the vector $\begin{pmatrix} \hat{a} - a \\ \hat{b} - b \end{pmatrix}$ is asymptotically distributed approximately as a bivariate normal random variable with mean $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and covariance matrix

$$\begin{pmatrix} \dfrac{1 - e^{-bt_k}}{a} & t_k e^{-bt_k} \\ t_k e^{-bt_k} & \displaystyle\sum_{i=1}^{k} \frac{a\left(t_i e^{-bt_i} - t_{i-1} e^{-bt_{i-1}}\right)^2}{\left(e^{-bt_{i-1}} - e^{-bt_i}\right)} \end{pmatrix}^{-1}$$

Performing the matrix inversion leads to the following:

$$\operatorname{Var}(\hat{a}) \stackrel{a}{=} \sum_{i=1}^{k} \frac{a\left(t_i e^{-bt_i} - t_{i-1} e^{-bt_{i-1}}\right)^2}{\left(e^{-bt_{i-1}} - e^{-bt_i}\right)} \Bigg/ D^*$$

$$\operatorname{Var}(\hat{b}) \stackrel{a}{=} \frac{\left(1 - e^{-bt_k}\right)}{a} \Bigg/ D^*$$

$$\operatorname{Cov}(\hat{a}, \hat{b}) \stackrel{a}{=} -t_k e^{-bt_k} \Big/ D^*$$

where

$$D^* = \frac{\left(1 - e^{-bt_k}\right)}{a} \sum_{i=1}^{k} \frac{a\left(t_i e^{-bt_i} - t_{i-1} e^{-bt_{i-1}}\right)^2}{\left(e^{-bt_{i-1}} - e^{-bt_i}\right)} - t_k^2 e^{-2bt_k}$$

To verify these results, the distributions of $\hat{a}$ and $\hat{b}$ were simulated by a similar procedure to that described in Section 4.4.1.1. The true value of a was 300 and that of b was 0.01. Again, k was taken to be 25 and $t_i = 10i$, $i = 1, 2, \ldots, 25$. Even though $k = 25$ is not large, the results derived here are consistent with the simulation results shown in Figure 4.4.1.2.

4.4.2 Large Sample Results: LSE

Since there is no literature available concerning the asymptotic properties for the least squares estimators a general procedure will be derived here which will be applied to the J-M and S-W and IBM models in Sections 4.4.2.1 and 4.4.2.2, respectively. Due to the complications existing in the Binomial models (e.g., successive observations are not independent) the asymptotic properties were not derived. For the three parameter GPM the algebraic manipulations and calculations are too unwieldly to be included here but may be carried out on a computer. The technique for three parameters is the same as that for two parameters which will now be illustrated.

Let $X_1, \ldots X_k, \ldots$ be independent random variables with finite moments $E(X_i) = g_i(\theta_1, \theta_2)$, $\mathrm{Var}(X_i) = \sigma_i^2(\theta_1, \theta_2)$ $E|X_i - g_i(\theta_1, \theta_2)|^3 = \rho_i(\theta_1, \theta_2)$ satisfying Liapunov's condition [c.f. (I,2)

$$(*) \quad \lim_{k \to \infty} \frac{\left\{ \sum_{i=1}^{k} \rho_i(\theta_1, \theta_2) \right\}^{1/3}}{\left\{ \sum_{i=1}^{k} \sigma_i^2(\theta_1, \theta_2) \right\}^{1/2}} = 0$$

for each finite value of the unknown parameters θ_1 and θ_2. It is assumed that $g_i(\theta_1, \theta_2)$ is, for each i, three times continuously differentiable in a neighborhood of the true value of $(\theta_1, \theta_2)^T$. Let

$$S(\theta_1, \theta_2) = \sum_{i=1}^{k} \left(X_i - g_i(\tilde{\theta}_1, \tilde{\theta}_2)\right)^2$$

and define $\tilde{\theta}_1$, $\tilde{\theta}_2$ (dependence on k has been suppressed for notational convenience) such that

$$S(\tilde{\theta}_1, \tilde{\theta}_2) = \sum_{i=1}^{k} \left(X_i - g_i(\tilde{\theta}_1, \tilde{\theta}_2)\right)^2 = \inf_{\{(\theta_1, \theta_2)\}} S(\theta_1, \theta_2).$$

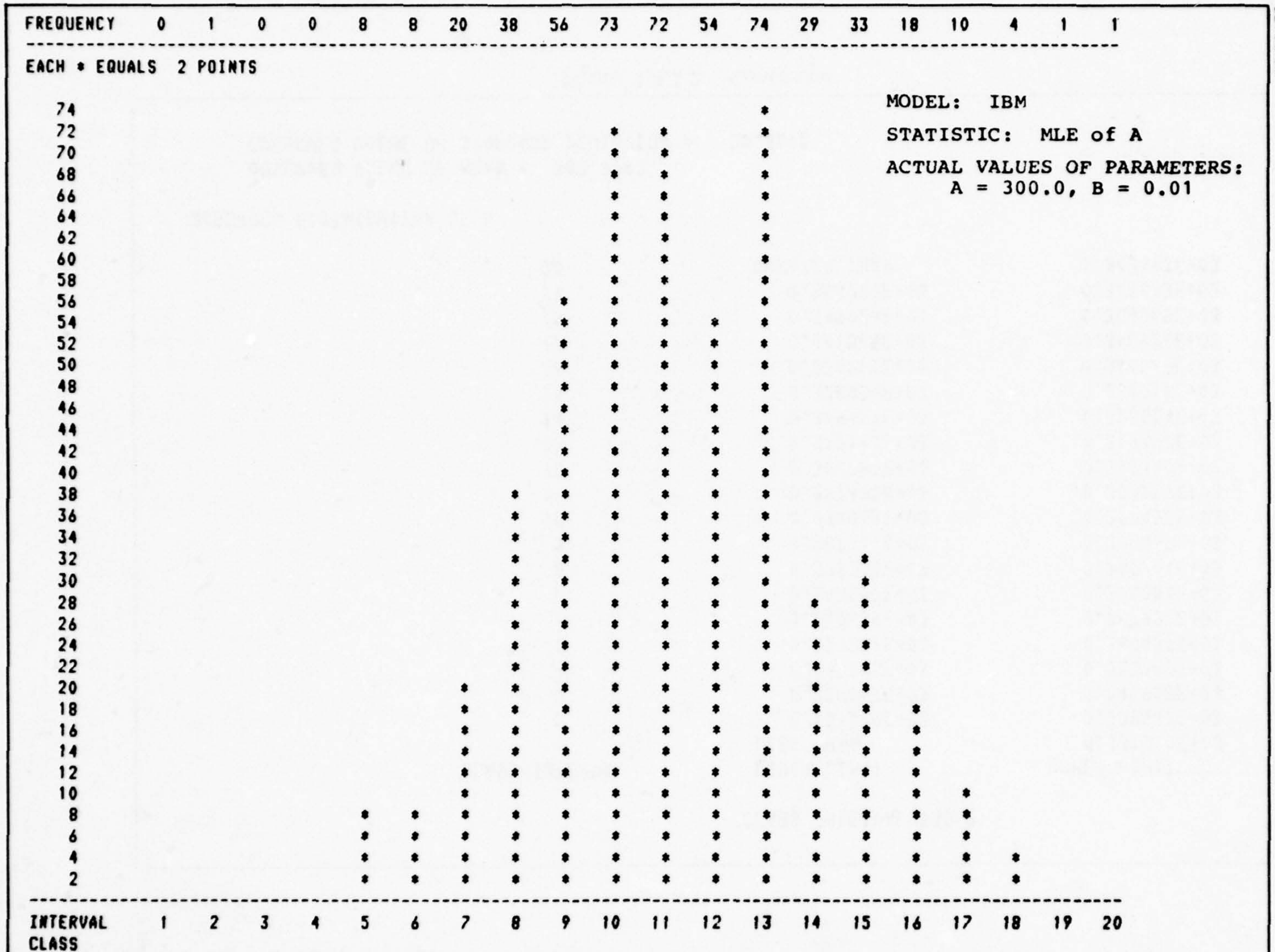

Figure 4.4.1.2. Simulation of the Distributions of the MLEs for the IBM Model

82755-14

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.231210E+03
2	0.231210E+03	0.238567E+03
3	0.238567E+03	0.245925E+03
4	0.245925E+03	0.253282E+03
5	0.253282E+03	0.260639E+03
6	0.260639E+03	0.267997E+03
7	0.267997E+03	0.275354E+03
8	0.275354E+03	0.282711E+03
9	0.282711E+03	0.290068E+03
10	0.290068E+03	0.297425E+03
11	0.297425E+03	0.304782E+03
12	0.304782E+03	0.312140E+03
13	0.312140E+03	0.319497E+03
14	0.319497E+03	0.326854E+03
15	0.326854E+03	0.334211E+03
16	0.334211E+03	0.341568E+03
17	0.341568E+03	0.348926E+03
18	0.348926E+03	0.356283E+03
19	0.356283E+03	0.363643E+03
20	GREATER THAN	0.363643E+03

OBSERVED DISTRIBUTION OF A

OBSERVED VALUE OF MEAN = 302.9792
OBSERVED VALUE OF STANDARD DEVIATION = 20.3695

Figure 4.4.1.2. Continued

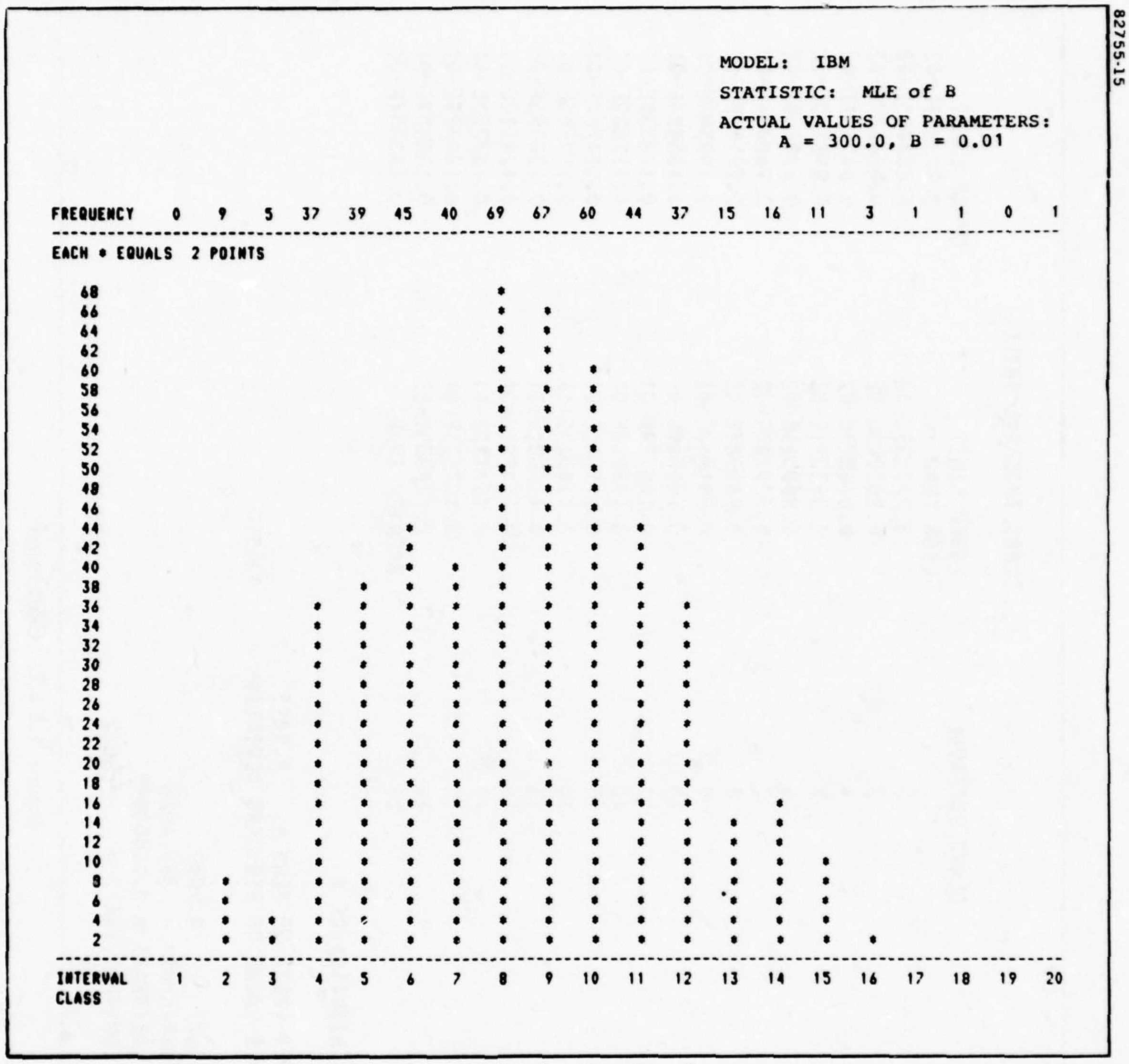

Figure 4.4.1.2. Continued

82755-16

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.773939E-02
2	0.773939E-02	0.805083E-02
3	0.805083E-02	0.836227E-02
4	0.836227E-02	0.867371E-02
5	0.867371E-02	0.898515E-02
6	0.898515E-02	0.929660E-02
7	0.929660E-02	0.960804E-02
8	0.960804E-02	0.991948E-02
9	0.991948E-02	0.102309E-01
10	0.102309E-01	0.105424E-01
11	0.105424E-01	0.108538E-01
12	0.108538E-01	0.111652E-01
13	0.111652E-01	0.114767E-01
14	0.114767E-01	0.117881E-01
15	0.117881E-01	0.120996E-01
16	0.120996E-01	0.124110E-01
17	0.124110E-01	0.127225E-01
18	0.127225E-01	0.130339E-01
19	0.130339E-01	0.133454E-01
20	GREATER THAN	0.133454E-01

OBSEVED DISTRIBUTION OF B

OBSERVED VALUE OF MEAN = 0.0099
OBSERVED VALUE OF STANDARD DEVIATION = 0.0010

SAMPLE COV(A,B) = -0.0080
ASYMPTOTIC VAR(AHAT) = 368.6828
ASYMPTOTIC VAR(BHAT) = 0.00000093
ASYMPTOTIC COV(AHAT,BHAT) = -0.0062

Figure 4.4.1.2. Continued

Because of the assumptions it is clear that $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are solutions to

$$\frac{\partial S}{\partial \theta_1} = -2 \sum_{i=1}^{k} (X_i - g_i(\theta_1, \theta_2)) \frac{\partial g_i}{\partial \theta_1} = 0$$

$$\frac{\partial S}{\partial \theta_2} = -2 \sum_{i=1}^{k} (X_i - g_i(\theta_1, \theta_2)) \frac{\partial g_i}{\partial \theta_2} = 0.$$

Because of (*) and under mild conditions on the first partial derivatives of g_i, $\begin{pmatrix} \partial S/\partial\theta_1 \\ \partial S/\partial\theta_2 \end{pmatrix}$ is asymptotically bivariate normally distributed with

mean $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and covariance matrix equal to

$$4 \cdot \begin{pmatrix} \sum_{i=1}^{k} \sigma_i^2(\theta_1, \theta_2) \left(\frac{\partial g_i}{\partial \theta_1}\right)^2 & \sum_{i=1}^{k} \sigma_i^2(\theta_1, \theta_2) \frac{\partial g_i}{\partial \theta_1} \frac{\partial g_i}{\partial \theta_2} \\ \sum_{i=1}^{k} \sigma_i^2(\theta_1, \theta_2) \frac{\partial g_i}{\partial \theta_1} \frac{\partial g_i}{\partial \theta_2} & \sum_{i=1}^{k} \sigma_i^2(\theta_1, \theta_2) \left(\frac{\partial g_i}{\partial \theta_2}\right)^2 \end{pmatrix}$$

$$= 4 \cdot \Sigma_1$$

when θ_1 and θ_2 are the true values of the parameters. Using a theorem from (I,1) it is seen that

$$\begin{pmatrix} \partial S/\partial\theta_1 \\ \partial S/\partial\theta_2 \end{pmatrix} \Bigg|_{\substack{\theta_1 = \tilde{\theta}_1 \\ \theta_2 = \tilde{\theta}_2}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} =$$

$$\begin{pmatrix} \partial S/\partial\theta_1 \\ \partial S/\partial\theta_2 \end{pmatrix} + \begin{pmatrix} \partial^2 S/\partial\theta_1^{\,2} & \partial^2 S/\partial\theta_1\partial\theta_2 \\ \partial^2 S/\partial\theta_1\partial\theta_2 & \partial^2 S/\partial\theta_2^2 \end{pmatrix} \cdot \begin{pmatrix} \tilde{\theta}_1 - \theta_1 \\ \tilde{\theta}_2 - \theta_2 \end{pmatrix} + R(\tilde{\theta}_1 - \tilde{\theta}_1, \tilde{\theta}_2 - \theta_2)$$

where $\left|R(\tilde{\theta}_1 - \tilde{\theta}_1, \tilde{\theta}_2 - \theta_2)\right| / \sqrt{(\tilde{\theta}_1 - \theta_1)^2 + (\tilde{\theta}_2 - \theta_2)^2}$ converges to zero in probability as $k \to \infty$. Solving for

$$\begin{pmatrix} \tilde{\theta}_1 - \theta_1 \\ \tilde{\theta}_2 - \theta_2 \end{pmatrix}$$

gives

$$\begin{pmatrix} \tilde{\theta}_1 - \theta_1 \\ \tilde{\theta}_2 - \theta_2 \end{pmatrix} \stackrel{a}{=} - \begin{pmatrix} \partial^2 S/\partial\theta_1^{\,2} & \partial^2 S/\partial\theta_1\partial\theta_2 \\ \partial^2 S/\partial\theta_1\partial\theta_2 & \partial^2 S/\partial\theta_2^{\,2} \end{pmatrix}^{-1} \cdot \begin{pmatrix} \partial S/\partial\theta_1 \\ \partial S/\partial\theta_2 \end{pmatrix}$$

$$= -\Sigma_2^{-1} \begin{pmatrix} \partial S/\partial\theta_1 \\ \partial S/\partial\theta_2 \end{pmatrix} .$$

Computing the derivatives in Σ_2 gives

$$\frac{\partial^2 S}{\partial\theta_1^{\,2}} = -2 \sum_{i=1}^{k} \left\{ (X_i - g_i(\theta_1, \theta_2)) \frac{\partial^2 g_i}{\partial\theta_1^{\,2}} - \left(\frac{\partial g_i}{\partial\theta_2}\right)^2 \right\}$$

$$\frac{\partial^2 S}{\partial\theta_2^{\,2}} = -2 \sum_{i=1}^{k} \left\{ (X_i - g_i(\theta_1, \theta_2)) \frac{\partial^2 g_i}{\partial\theta_2^{\,2}} - \left(\frac{\partial g_i}{\partial\theta_2}\right)^2 \right\}$$

$$\frac{\partial^2 S}{\partial\theta_1\partial\theta_2} = -2 \sum_{i=1}^{k} \left\{ (X_i - g_i(\theta_1, \theta_2)) \frac{\partial^2 g_i}{\partial\theta_1\partial\theta_2} - \frac{\partial g_i}{\partial\theta_1} \frac{\partial g_i}{\partial\theta_2} \right\} .$$

Under the additional assumptions that as $k \to \infty$,

$$\frac{1}{k^2} \sum_{i=1}^{k} \left(\frac{\partial^2 g_i}{\partial \theta_1^2} \right)^2 \sigma_i^2 (\theta_1, \theta_2) \to 0$$

$$\frac{1}{k^2} \sum_{i=1}^{k} \left(\frac{\partial^2 g_i}{\partial \theta_2^2} \right)^2 \sigma_i^2 (\theta_1, \theta_2) \to 0$$

$$\frac{1}{k^2} \sum_{i=1}^{k} \left(\frac{\partial^2 g_i}{\partial \theta_1 \partial \theta_2} \right)^2 \sigma_i^2 (\theta_1, \theta_2) \to 0$$

The matrix $\sum_2$ is asymptotically equal to

$$2 \cdot \begin{pmatrix} \sum_{i=1}^{k} \left(\frac{\partial g_i}{\partial \theta_1} \right)^2 & \sum_{i=1}^{k} \frac{\partial g_i}{\partial \theta_1} \frac{\partial g_i}{\partial \theta_2} \\ \sum_{i=1}^{k} \frac{\partial g_i}{\partial \theta_1} \frac{\partial g_i}{\partial \theta_2} & \sum_{i=1}^{k} \left(\frac{\partial g_i}{\partial \theta_2} \right)^2 \end{pmatrix}.$$

in probability. The matrix $\sum_2^{-1}$ is thus asymptotically equal (in probability) to

$$\frac{1}{2D'} \cdot \begin{pmatrix} \sum_{i=1}^{k} \left(\frac{\partial g_i}{\partial \theta_2} \right)^2 & - \sum_{i=1}^{k} \frac{\partial g_i}{\partial \theta_1} \frac{\partial g_i}{\partial \theta_2} \\ -\sum_{i=1}^{k} \frac{\partial g_i}{\partial \theta_1} \frac{\partial g_i}{\partial \theta_2} & \sum_{i=1}^{k} \left(\frac{\partial g_i}{\partial \theta_1} \right)^2 \end{pmatrix} = \frac{1}{2D'} \sum_3$$

where

$$D' = \sum_{i=1}^{k} \left(\frac{\partial g_i}{\partial \theta_1}\right)^2 \sum_{i=1}^{k} \left(\frac{\partial g_i}{\partial \theta_2}\right)^2 - \left(\sum_{i=1}^{k} \frac{\partial g_i}{\partial \theta_1} \frac{\partial g_i}{\partial \theta_2}\right)^2 .$$

The result is thus that

$$\begin{pmatrix} \tilde{\theta}_1 - \theta_1 \\ \tilde{\theta}_2 - \theta_2 \end{pmatrix}$$

is asymptotically normally distributed with mean $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and covariance matrix

$$\frac{1}{(D')^2} \quad \Sigma_3 \cdot \begin{pmatrix} \sum_{i=1}^{k} \sigma_i^2 (\theta_1, \theta_2) \left(\frac{\partial g_i}{\partial \theta_1}\right)^2 & \sum_{i=1}^{k} \sigma_i^2 (\theta_1, \theta_2) \frac{\partial g_i}{\partial \theta_1} \frac{\partial g_i}{\partial \theta_2} \\ \sum_{i=1}^{k} \sigma_i^2 (\theta_1, \theta_2) \frac{\partial g_i}{\partial \theta_1} \frac{\partial g_i}{\partial \theta_2} & \sum_{i=1}^{k} \sigma_i^2 (\theta_1, \theta_2) \left(\frac{\partial g_i}{\partial \theta_2}\right)^2 \end{pmatrix} \cdot \Sigma_3 .$$

4.4.2.1 The GPM, J-M, and S-W Models

The general discussion in Section 4.4.2 can now be applied to the J-M and S-W models. As mentioned earlier the asymptotic covariance matrix for the estimators in the GPM may be derived in a manner similar to that illustrated in 4.4.2 and the computations carrier out on a computer.

For the J-M ($\alpha = 1$) or S-W ($\alpha = 2$) models, θ_1 and θ_2 in Section 4.4.2 may be taken to be N and ϕ, respectively and

$$g_i(\theta_1, \theta_2) = g_i(N, \phi) = \phi(N - M_{i-1})\tau_i^{\alpha}$$

$$\sigma_i^2(\theta_1, \theta_2) = \sigma_i^2(N, \phi) = \phi(N - M_{i-1})\tau_i^{\alpha}$$

for i = 1, 2, ... , k, ... Computing the necessary partial derivatives leads to (see Section 4.4.2 for notation and symbols)

$$\Sigma_3 = \begin{pmatrix} \sum_{i=1}^{k} (N - M_{i-1})^2 \tau_i^{2\alpha} & - \sum_{i=1}^{k} \phi(N-M_{i-1})\tau_i^{2\alpha} \\ - \sum_{i=1}^{k} \phi(N - M_{i-1})\tau_i^{2\alpha} & \sum_{i=1}^{k} \phi^2 \tau_i^{2\alpha} \end{pmatrix}$$

$$D' = \left(\sum_{i=1}^{k} \phi^2 \tau_i^{2\alpha} \sum_{i=1}^{k} (N - M_{i-1})^2 \tau_i^{2\alpha} - \left(\sum_{i=1}^{k} \phi(N - M_{i-1}) \tau_i^{2\alpha} \right)^2 \right)$$

So that $\begin{pmatrix} \tilde{N} - N \\ \tilde{\phi} - \phi \end{pmatrix}$ is asymptotically normally distributed with mean $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and

convariance matrix

$$\frac{1}{(D')^2} \Sigma_3 \begin{pmatrix} \sum_{i=1}^{k} \phi^3 (N - M_{i-1})\tau_i^{3\alpha} & \sum_{i=1}^{k} \phi^2 (N - M_{i-1})^2 \tau_i^{3\alpha} \\ \sum_{i=1}^{k} \phi^2 (N - M_{i-1})^2 \tau_i^{3\alpha} & \sum_{i=1}^{k} \phi^3 (N - M_{i-1})\tau_i^{3\alpha} \end{pmatrix} \cdot \Sigma_3$$

As usual, for the J-M model replace α by 1 and for the S-W model, replace α by 2. To verify these results, the sampling distributions of $\tilde{N}$ and $\tilde{\phi}$ were simulated for the J-M and S-W models under the same conditions as described in Section 4.4.1.1, i.e., actual value of N equal to 300, actual ϕ of 0.01, etc. The results are shown in Figure 4.4.2.1. Note how closely the asymptotic results match for k = 25.

4.4.2.2 IBM Model

To apply the results of Section 4.4.2 to the IBM model let $\theta_1 = a$, $\theta_2 = b$, and

$$g_i(\theta_1, \theta_2) = g_i(a, b) = a\left(e^{-bt_{i-1}} - e^{-bt_i}\right)$$

$$\sigma_i^2(\theta_1, \theta_2) = \sigma_i^2(a, b) = a\left(e^{-bt_{i-1}} - e^{-bt_i}\right)$$

where

$$t_0 \equiv 0, \quad i = 1, 2, \cdots, k, \cdots.$$

Taking the necessary partial derivatives and referring to Section 4.4.2 for notation and symbols, it is seen that

$$\sum_3 = \begin{pmatrix} \sum_{i=1}^{k} \left[a\left(t_i e^{-bt_i} - t_{i-1} e^{-bt_{i-1}}\right)\right]^2 & -\sum_{i=1}^{k} a\left(t_i e^{-bt_i} - t_{i-1} e^{-bt_{i-1}}\right)\left(e^{-bt_{i-1}} - e^{-bt_i}\right) \\ -\sum_{i=1}^{k} a\left(t_i e^{-bt_i} - t_{i-1} e^{-bt_{i-1}}\right)\left(e^{-bt_{i-1}} - e^{-bt_i}\right) & \sum_{i=1}^{k}\left(e^{-bt_{i-1}} - e^{-bt_i}\right)^2 \end{pmatrix}$$

and

$$D' = \sum_{i=1}^{k}\left[a\left(t_i e^{-bt_i} - t_{i-1} e^{-bt_{i-1}}\right)\right]^2 \sum_{i=1}^{k}\left(e^{-bt_{i-1}} - e^{-bt_i}\right)^2 -$$

$$\left(\sum_{i=1}^{k} a\left(t_i e^{-bt_i} - t_{i-1} e^{-bt_{i-1}}\right)\left(e^{bt_{i-1}} - e^{-bt_i}\right)\right)^2$$

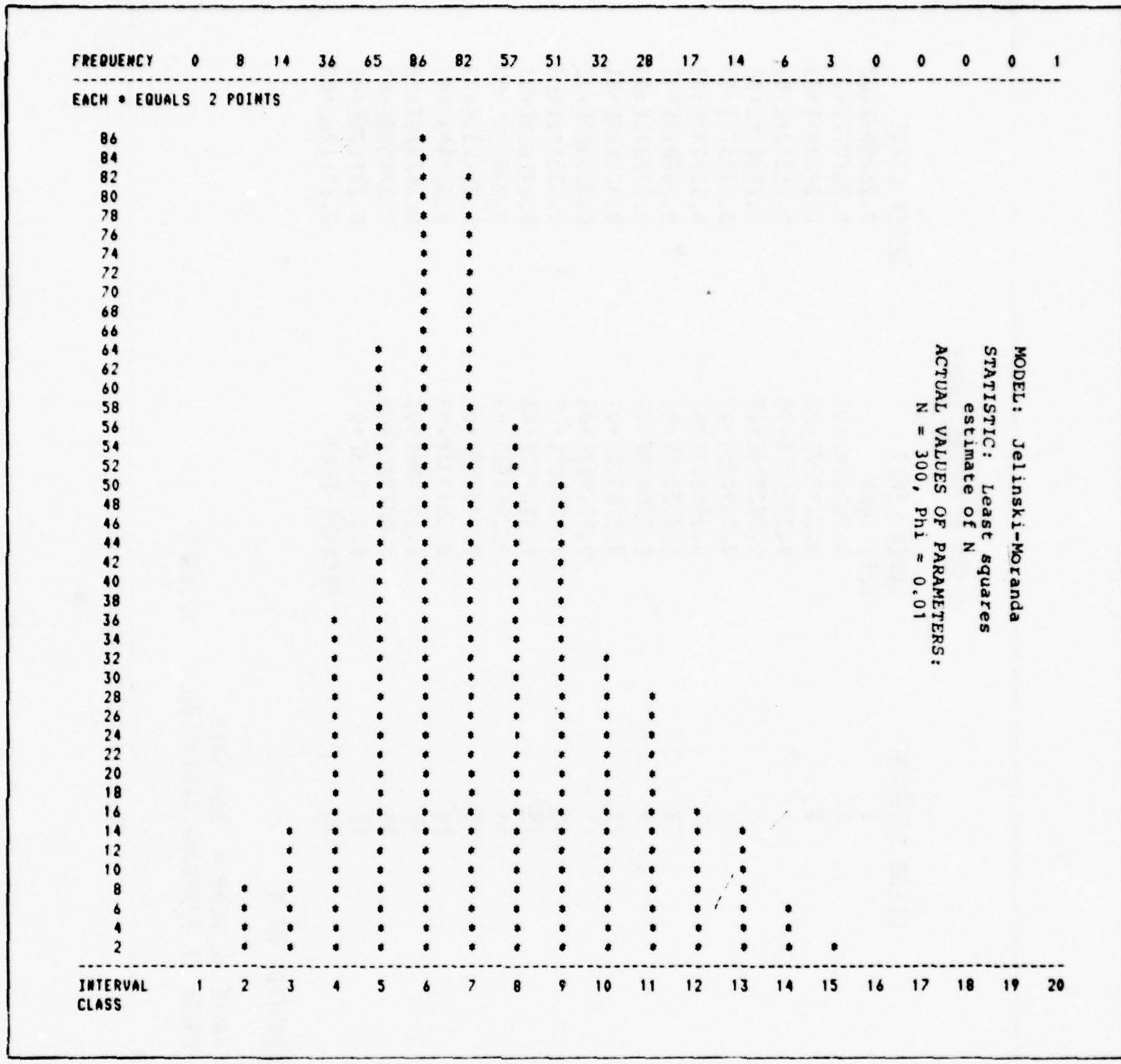

Figure 4.4.2.1. Simulation of the Distributions of the LSEs for the J-M and S-W Models

82755-18

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.262804E+03
2	0.262804E+03	0.269377E+03
3	0.269377E+03	0.275951E+03
4	0.275951E+03	0.282524E+03
5	0.282524E+03	0.289098E+03
6	0.289098E+03	0.295671E+03
7	0.295671E+03	0.302245E+03
8	0.302245E+03	0.308818E+03
9	0.308818E+03	0.315392E+03
10	0.315392E+03	0.321965E+03
11	0.321965E+03	0.328539E+03
12	0.328539E+03	0.335112E+03
13	0.335112E+03	0.341686E+03
14	0.341686E+03	0.348259E+03
15	0.348259E+03	0.354833E+03
16	0.354833E+03	0.361406E+03
17	0.361406E+03	0.367980E+03
18	0.367980E+03	0.374553E+03
19	0.3[illegible]E+03	0.381128E+03
20	GREAT[illegible] THAN	0.381128E+03

OBSERVED DISTRIBUTION OF N

OBSERVED VALUE OF MEAN = 301.5396
OBSERVED VALUE OF STANDARD DEVIATION = 17.6305

Figure 4.4.2.1. Continued

82755-19

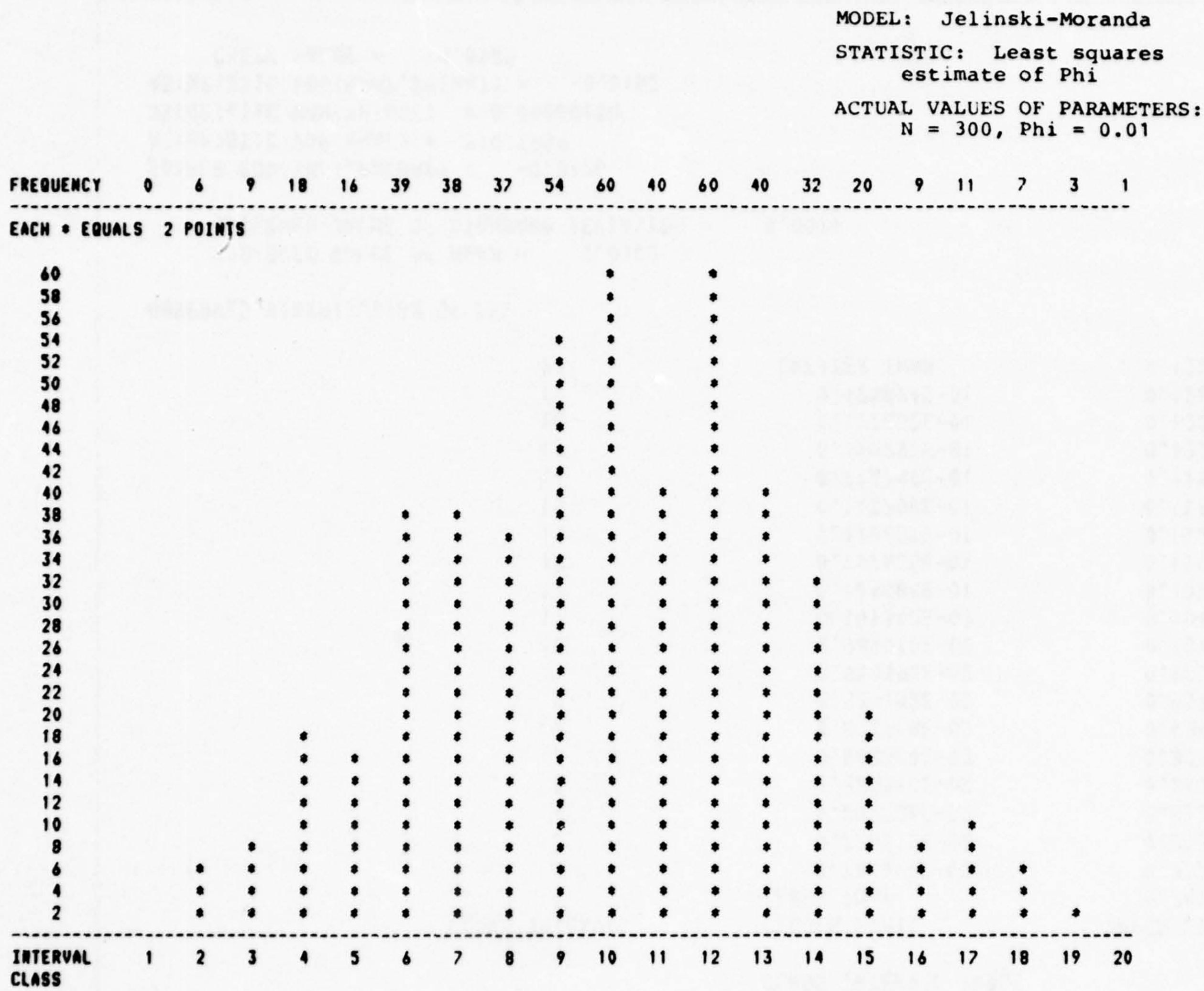

Figure 4.4.2.1. Continued

82755-20

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.741698E-02
2	0.741698E-02	0.772112E-02
3	0.772112E-02	0.802526E-02
4	0.802526E-02	0.832940E-02
5	0.832940E-02	0.863354E-02
6	0.863354E-02	0.893768E-02
7	0.893768E-02	0.924182E-02
8	0.924182E-02	0.954596E-02
9	0.954596E-02	0.985010E-02
10	0.985010E-02	0.101542E-01
11	0.101542E-01	0.104584E-01
12	0.104584E-01	0.107625E-01
13	0.107625E-01	0.110667E-01
14	0.110667E-01	0.113708E-01
15	0.113708E-01	0.116749E-01
16	0.116749E-01	0.119791E-01
17	0.119791E-01	0.122832E-01
18	0.122832E-01	0.125874E-01
19	0.125874E-01	0.128916E-01
20	GREATER THAN	0.128916E-01

OBSERVED DISTRIBUTION OF PHI

OBSERVED VALUE OF MEAN = 0.0100
OBSERVED VALUE OF STANDARD DEVIATION = 0.0011

SAMPLE COV(NHAT,PHIHAT) = -0.0170
ASYMPTOTIC VAR(NHAT) = 310.3989
ASYMPTOTIC VAR(PHIHAT) = 0.00000130
ASYMPTOTIC COV(NHAT,PHIHAT) = -0.0182
CHECK VALUE = -0.0182

Figure 4.4.2.1. Continued

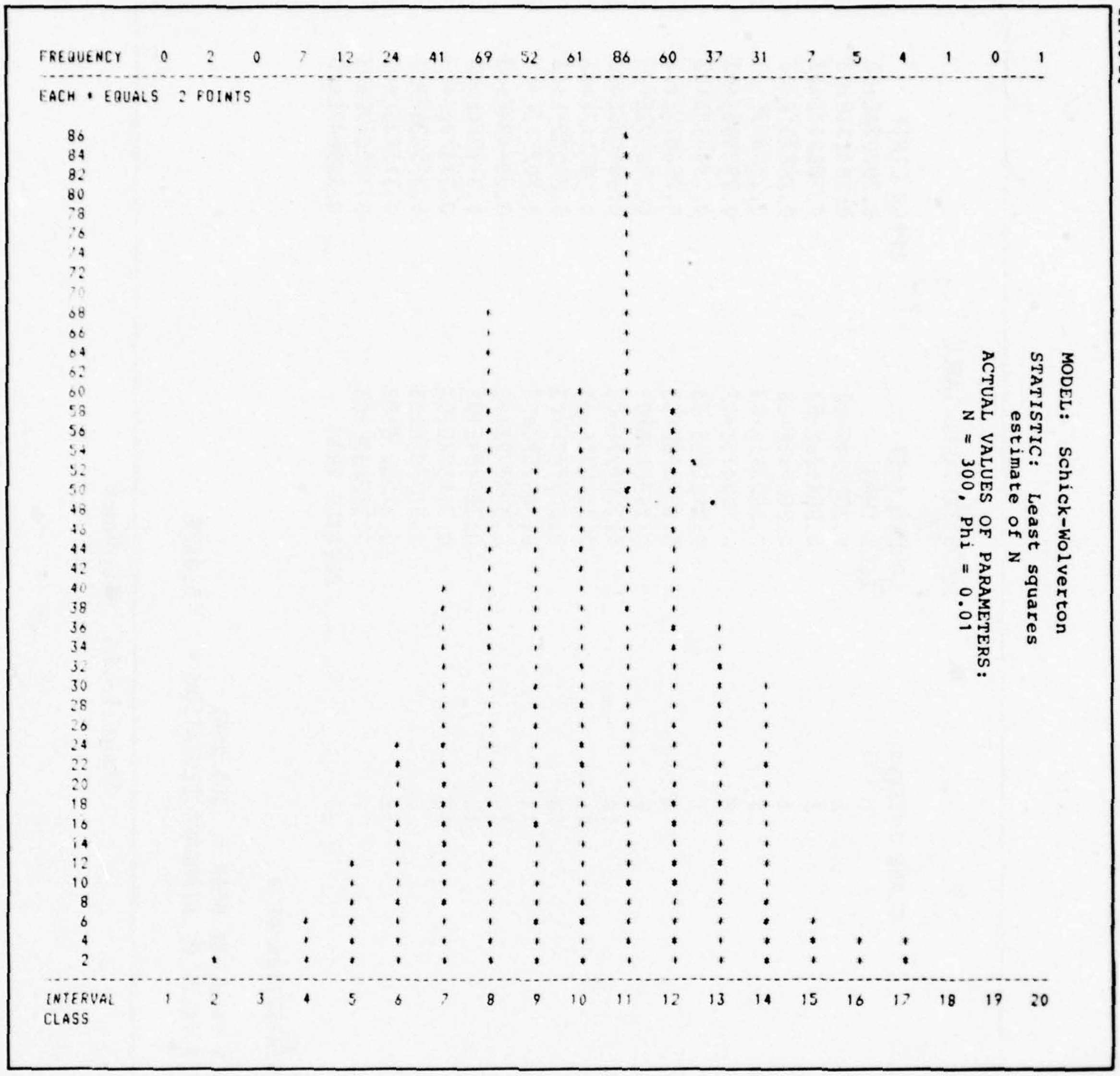

Figure 4.4.2.1. Continued

82755-22

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.282024E+03
2	0.282024E+03	0.284136E+03
3	0.284136E+03	0.286249E+03
4	0.286249E+03	0.288361E+03
5	0.288361E+03	0.290473E+03
6	0.290473E+03	0.292585E+03
7	0.292585E+03	0.294698E+03
8	0.294698E+03	0.296810E+03
9	0.296810E+03	0.298922E+03
10	0.298922E+03	0.301035E+03
11	0.301035E+03	0.303147E+03
12	0.303147E+03	0.305259E+03
13	0.305259E+03	0.307372E+03
14	0.307372E+03	0.309484E+03
15	0.309484E+03	0.311596E+03
16	0.311596E+03	0.313708E+03
17	0.313708E+03	0.315821E+03
18	0.315821E+03	0.317933E+03
19	0.317933E+03	0.320047E+03
20	GREATER THAN	0.320047E+03

OBSERVED DISTRIBUTION OF N

OBSERVED VALUE OF MEAN = 300.0542
OBSERVED VALUE OF STANDARD DEVIATION = 5.6475

Figure 4.4.2.1. Continued

82755-23

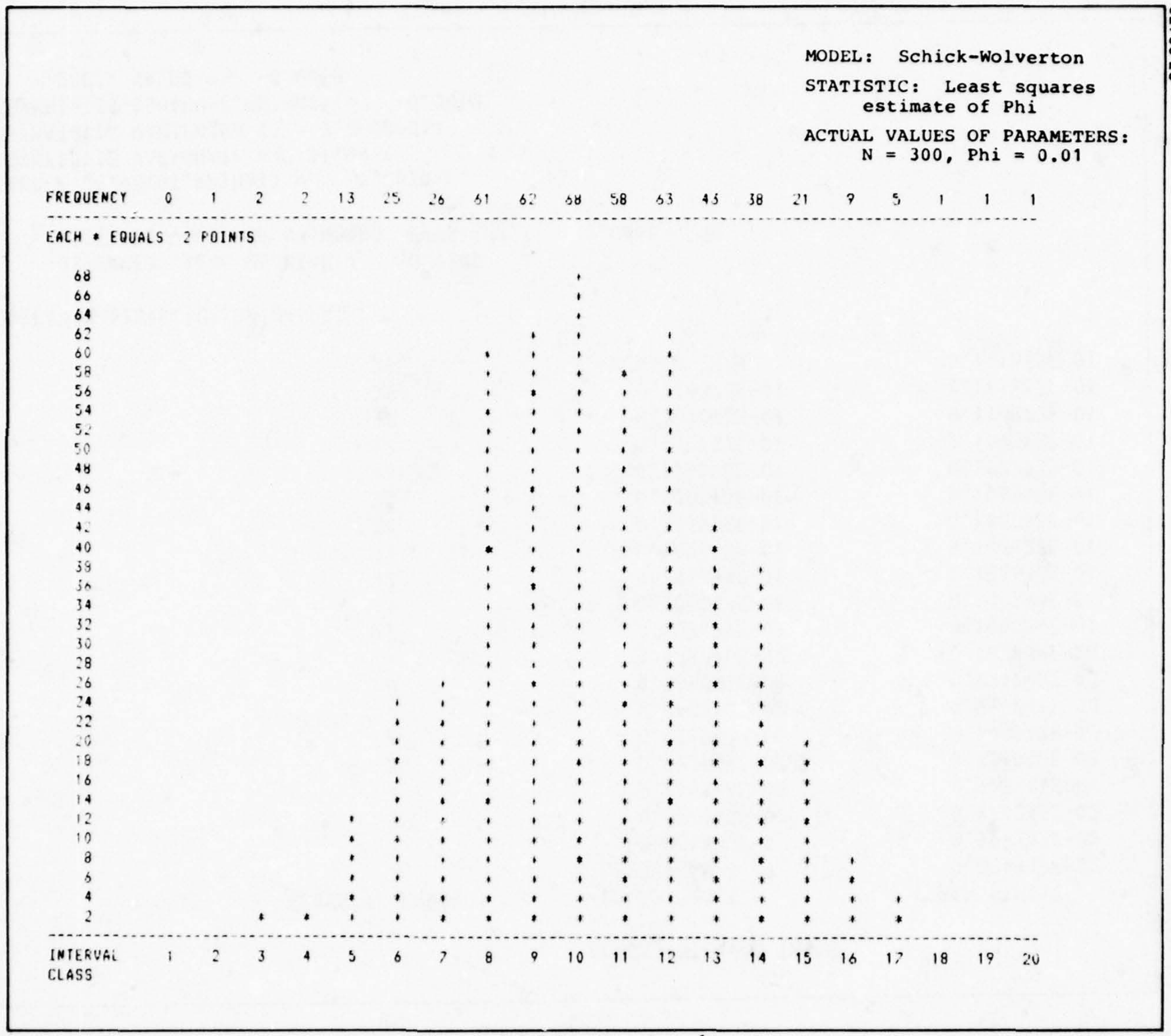

Figure 4.4.2.1. Continued

82755-24

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.884448E-02
2	0.884448E-02	0.897347E-02
3	0.897347E-02	0.910246E-02
4	0.910246E-02	0.923144E-02
5	0.923144E-02	0.936043E-02
6	0.936043E-02	0.948942E-02
7	0.948942E-02	0.961841E-02
8	0.961841E-02	0.974740E-02
9	0.974740E-02	0.987639E-02
10	0.987639E-02	0.100054E-01
11	0.100054E-01	0.101344E-01
12	0.101344E-01	0.102633E-01
13	0.102633E-01	0.103923E-01
14	0.103923E-01	0.105213E-01
15	0.105213E-01	0.106503E-01
16	0.106503E-01	0.107793E-01
17	0.107793E-01	0.109083E-01
18	0.109083E-01	0.110373E-01
19	0.110373E-01	0.111663E-01
20	GREATER THAN	0.111663E-01

OBSERVED DISTRIBUTION OF PHI

OBSERVED VALUE OF MEAN = 0.0100
OBSERVED VALUE OF STANDARD DEVIATION = 0.0004

SAMPLE COV(NHAT,PHIHAT) = -0.0020
ASYMPTOTIC VAR(NHAT) = 31.0411
ASYMPTOTIC VAR(PHIHA T) = 0.00000013
ASYMPTOTIC COV(NHAT,PHIHAT) = -0.0018
CHECK VALUE = -0.0018

Figure 4.4.2.1. Continued

so that $\begin{pmatrix} \tilde{a} - a \\ \tilde{b} - b \end{pmatrix}$ is asymptotically normally distributed with mean $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and covariance matrix

$$\frac{1}{(D')^2} \Sigma_3 \begin{pmatrix} \sum_{i=1}^{k} a\left(e^{-bt_{i-1}} - e^{-bt_i}\right)^3 & \sum_{i=1}^{k} a^2\left(e^{-bt_{i-1}} - e^{-bt_i}\right)^2\left(t_i e^{-bt_i} - t_{i-1}e^{-bt_{i-1}}\right) \\ \sum_{i=1}^{k} a^2\left(e^{-bt_{i-1}} - e^{-bt_i}\right)^2\left(t_i e^{-bt_i} - t_{i-1}e^{-bt_{i-1}}\right) & \sum_{i=1}^{k} a^3\left(e^{-bt_{i-1}} - e^{-bt_i}\right)\left(t_i e^{-bt_i} - t_{i-1}e^{-bt_{i-1}}\right)^2 \end{pmatrix} \Sigma_3$$

To verify these results, the sampling distribution of $\tilde{a}$ and $\tilde{b}$ were simulated under the same conditions described in Section 4.4.1.2, i.e., true a of 300, true b of 0.01, etc. The results are shown in Figure 4.4.2.2. Note, as in the case of the J-M and S-W models, the asymptotic results are very accurate for $k = 25$.

82755-25

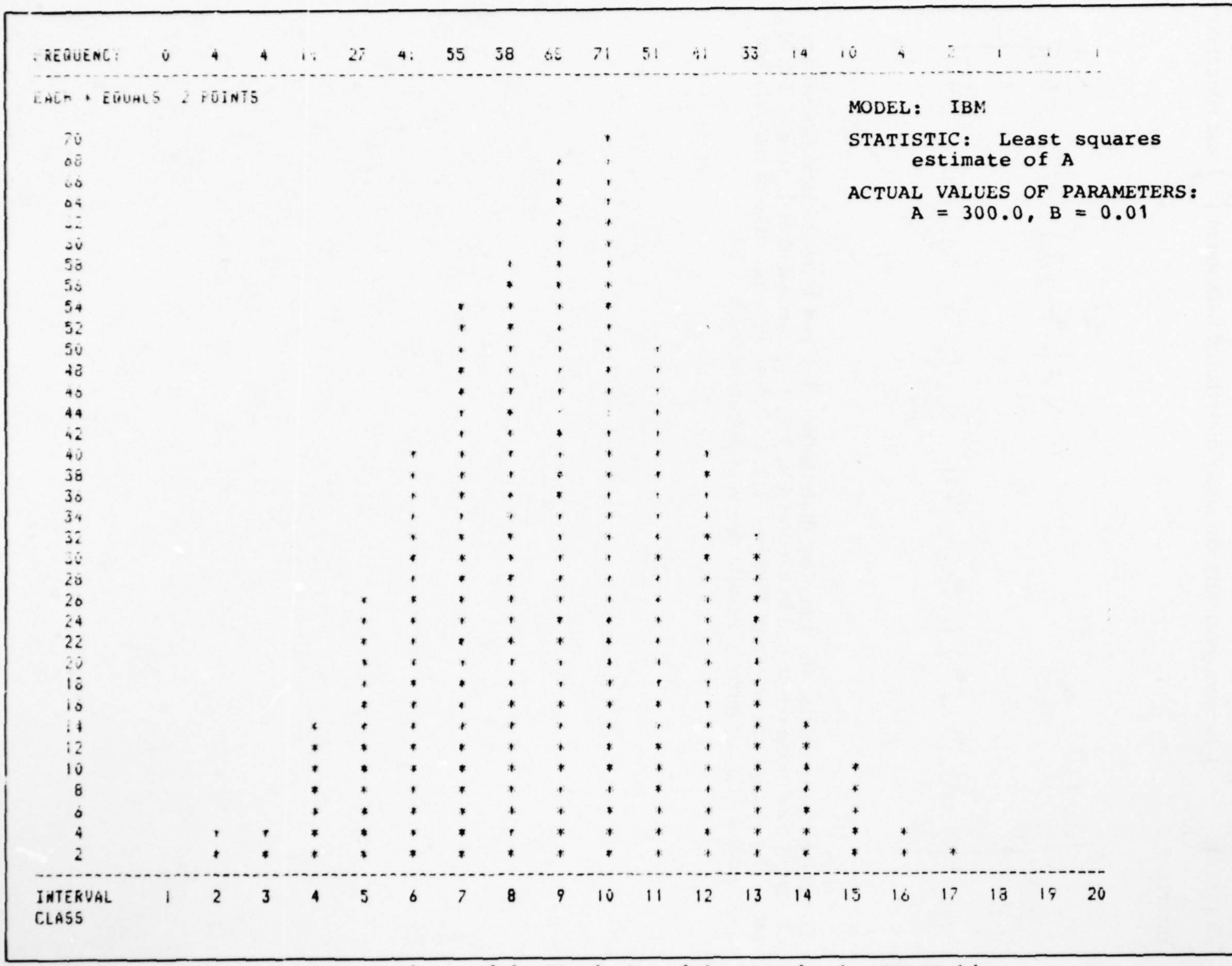

Figure 4.4.2.2. Simulation of the Distribution of the LSEs for the IBM Model

82755-26

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.244890E+03
2	0.244890E+03	0.252229E+03
3	0.252229E+03	0.259568E+03
4	0.259568E+03	0.266908E+03
5	0.266908E+03	0.274247E+03
6	0.274247E+03	0.281586E+03
7	0.281586E+03	0.288926E+03
8	0.288926E+03	0.296265E+03
9	0.296265E+03	0.303604E+03
10	0.303604E+03	0.310944E+03
11	0.310944E+03	0.318283E+03
12	0.318283E+03	0.325623E+03
13	0.325623E+03	0.332962E+03
14	0.332962E+03	0.340301E+03
15	0.340301E+03	0.347641E+03
16	0.347641E+03	0.354980E+03
17	0.354980E+03	0.362319E+03
18	0.362319E+03	0.369659E+03
19	0.369659E+03	0.377000E+03
20	GREATER THAN	0.377000E+03

OBSERVED DISTRIBUTION OF A

OBSERVED VALUE OF MEAN = 301.4595
OBSERVED VALUE OF STANDARD DEVIATION = 21.2965

Figure 4.4.2.2. Continued

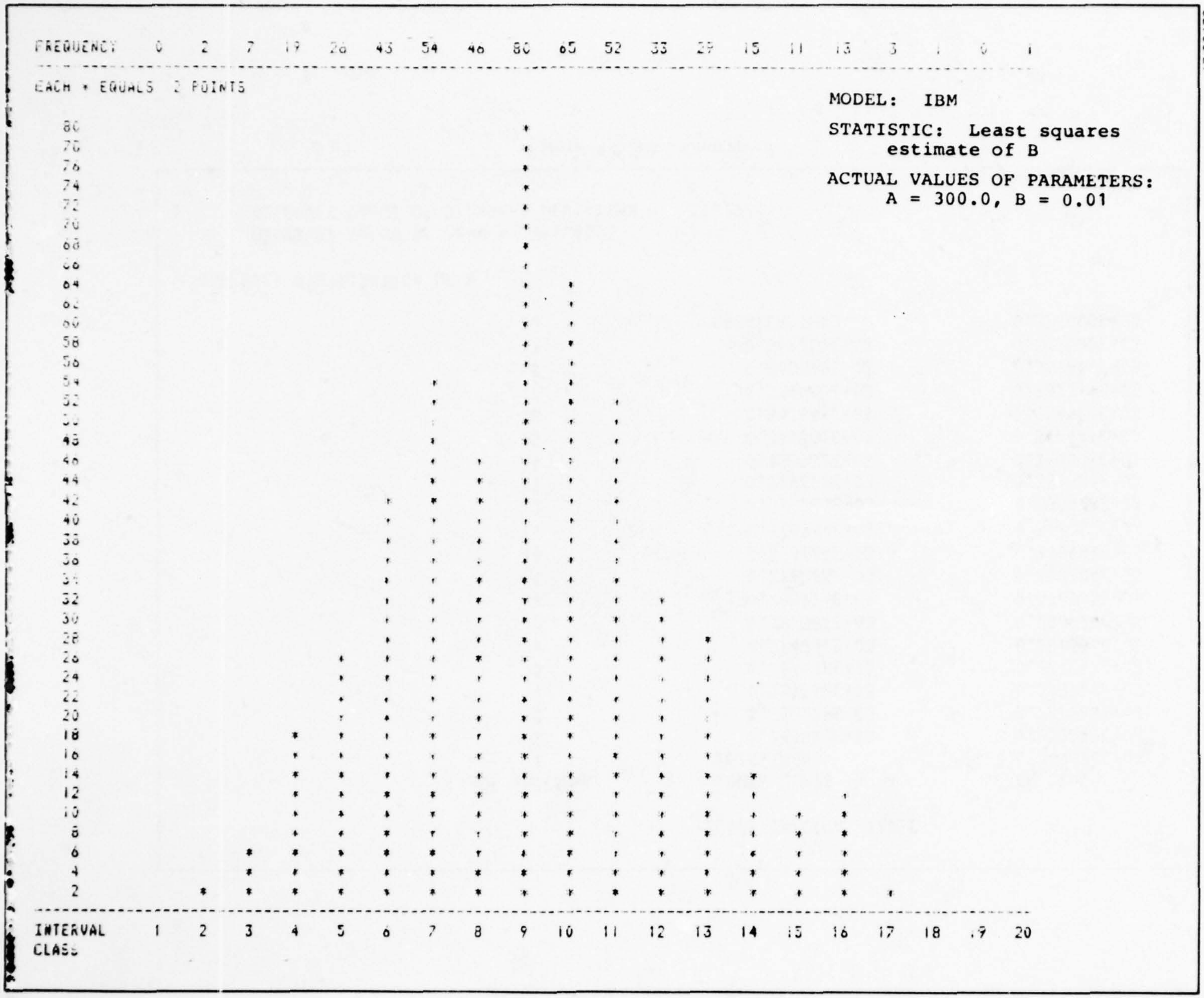

Figure 4.4.2.2. Continued

82755-28

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.712839E-02
2	0.712839E-02	0.750384E-02
3	0.750384E-02	0.787930E-02
4	0.787930E-02	0.825476E-02
5	0.825476E-02	0.863022E-02
6	0.863022E-02	0.900567E-02
7	0.900567E-02	0.938113E-02
8	0.938113E-02	0.975659E-02
9	0.975659E-02	0.101320E-01
10	0.101320E-01	0.105075E-01
11	0.105075E-01	0.108830E-01
12	0.108830E-01	0.112584E-01
13	0.112584E-01	0.116339E-01
14	0.116339E-01	0.120093E-01
15	0.120093E-01	0.123848E-01
16	0.123848E-01	0.127602E-01
17	0.127602E-01	0.131357E-01
18	0.131357E-01	0.135112E-01
19	0.135112E-01	0.138866E-01
20	GREATER THAN	0.138866E-01

OBSERVED DISTRIBUTION OF B

OBSERVED VALUE OF MEAN = 0.0100
OBSERVED VALUE OF STANDARD DEVIATION = 0.0012

SAMPLE COV(A,B) = -0.0116
ASYMPTOTIC VAR(AHAT) = 128.0872
ASYMPTOTIC VAR(BHAT) = 0.00000137
ASYMPTOTIC COV(AHAT,BHAT) = -0.0113
CHECK VALUE = -0.0113

Figure 4.4.2.2. Continued

4.4.3 Small Sample Considerations

Due to the nature of the models considered here, the typical statistical techniques used in deriving small sample distributions and confidence bounds on unknown parameters failed and further investigation yielded no significant results. Furthermore, one may be tempted to use the asymptotic results to derive confidence bounds since they were shown to be quite good for k as small as 25. However, it should be noted that for values of the parameters other than those used in the simulations $k = 25$ may not be sufficient.

4.5 ESTIMATING THE TIME REQUIRED TO REMOVE A SPECIFIED NUMBER OF THE REMAINING ERRORS

4.5.1 Introduction

Each of the models contains parameters which, in order to use the models for predicting reliability behavior, must be estimated. From these parameter estimates other more important parameters can be estimated. In this section we treat the estimation of expected time to remove a specified number of the remaining errors for the GPM, J-M, S-W, binomial, and IBM models. It will be assumed that M errors have been removed "to date" and that each remaining error is removed when it is observed. That is, what will be of concern are the times between successive error finds (removals) denoted by $S_{M+1}, S_{M+2} \cdots$. For the GPM, J-M, S-W, and binomial models, the sequence will terminate with S_N, the time between the N-1 st and Nth errors (by definition, the Nth error is the last error for these models).

4.5.2 GPM, J-M, and S-W Models

In this section the analysis will be carried out for the GPM model since the J-M and S-W results will follow by $\alpha=1$ and $\alpha=2$, respectively, in the results. Under the assumptions of Section 4.5.1, the expected time to remove $k \le N-M$ of the remaining N-M errors is

$$T_k = E\left(\sum_{i=M+1}^{M+k} S_i\right) = \sum_{i=M+1}^{M+k} E(S_i) \text{ , } k = 1, 2, \ldots, N-M.$$

The distributions of the S_i's are easily obtained as follows: The event $\{S_i \ge t\}$ (for $t > 0$) is equivalent to the event {number of errors in time interval of length t beginning at the time when the i-1st error occurred is zero}. Since this interval has length t and a total of i-1 errors have been removed up to the time of the i-1st error find/removal, it is seen that

$$P\{S_i \ge t\} = \frac{\{\phi(N-(i-1))\,t^{\alpha}\}^0}{0!} e^{-\phi(N-(i-1))\,t^{\alpha}}$$

$$= e^{-\phi(N-i+1)t^{\alpha}} \text{ , } t > 0.$$

The probability density function of S_i is

$$f_{S_i}(t) = \alpha\phi(N-i+1)t^{\alpha-1} e^{-\phi(N-i+1)t^{\alpha}}, \quad t>0;$$

a Weibull distribution.

The expectation of S_i is

$$E(S_i) = \int_0^{\infty} t f_{S_i}(t)\, dt$$

$$= \{\phi(N-i+1)\}^{-\frac{1}{\alpha}} \alpha^{-1} \Gamma(\alpha^{-1})$$

(noting that $\Gamma(\cdot)$ is the usual gamma function) and hence,

$$T_k = \alpha^{-1}\Gamma(\alpha^{-1}) \sum_{i=M+1}^{M+k} \{\phi(N-i+1)\}^{-\frac{1}{\alpha}}$$

Since N, ϕ, α are unknown in general, two estimates of T_k, $\hat{T}_k$ and $\tilde{T}_k$ are available by replacing N, ϕ, and α by their MLEs or LSEs, respectively. For example,

$$\hat{T}_k = \hat{\alpha}^{-1}\Gamma(\hat{\alpha}^{-1}) \sum_{i=M+1}^{M+k} \{\hat{\phi}(\hat{N}-i+1)\}^{-\frac{1}{\hat{\alpha}}}.$$

It should be pointed out that when estimates are used for N, ϕ, α, k must be less than or equal to $\hat{N}$-M or $\tilde{N}$-M, depending on the method used (MLE or LSE).

4.5.3 Binomial Models

The expression for the expected time to remove $k \leq N-M$ of the remaining N-M errors is , as in Section 4.5.2,

$$T_k = \sum_{i=M+1}^{M+k} E(S_i), \quad k = 1, 2, \ldots, N-M.$$

The probability density function of S_i is derived as follows: The event (for $t > 0$

$\{S_i \geq t\}$ is equal to the event

{no. errors in the time interval of length t beginning at the time of the i-1[st] error find/removal is zero}. Using the Binomial model (see Section 3.3 for notations),

$$P\{S_i \geq t\} = \binom{N-(i-1)}{0} p(t)^0 (1-p(t))^{N-(i-1)-0}$$

$$= (1-p(t))^{N-i+1}, \quad t > 0.$$

For Binomial I, $p(t) = t/(t+c)$. In this case,

$$P\{S_i \geq t\} = \{c/(t+c)\}^{N-i+1}, \quad t > 0,$$

and the probability density function of S_i is

$$f_{S_i}(t) = \{N-i+1\} \{c/(t+c)\}^{N-i} c/(t+c)^2$$

$$= \{N-i+1\} c^{(N-i+1)} \Big/ (t+c)^{(N-i+2)}, \quad t > 0.$$

The expected value of S_i is

$$E(S_i) = \int_0^\infty t f_{S_i}(t)\, dt$$

$$= \begin{cases} \infty & \text{when } i = N \\ & \text{(i.e. the last error is never found,} \\ & \text{on the average)} \\ c\,(N-i)^{-1} & \text{for } i = M+1, M+2, \ldots, M+k;\ k = 1, \ldots, N-M-1. \end{cases}$$

For the Binomial II, $p(t) = 1-e^{-at}$ so that

$$P\{S_i \geq t\} = e^{-a(N-i+1)t}$$

and the probability density function of S_i is

$$f_{S_i}(t) = a(N-i+1)\, e^{-a(N-i+1)t}$$

with

$$E(S_i) = \{a(N-i+1)\}^{-1} .$$

Due to the one case of infinite expection for the Binomial I model, T_k is not defined when $k = N-M$. Thus, for the Binomial I model,

$$T_k = \sum_{i=M+1}^{M+k} c\,(N-i)^{-1}, \quad k = 1, 2, \ldots, N-M-1,$$

with N, c replaced by $\hat{N}$, $\hat{c}$ or $\tilde{N}$, $\tilde{c}$ for estimators. For the Binomial II model.

$$T_k = \sum_{i=M+1}^{M+k} \{a(N-i+1)\}^{-1}, \quad k = 1, 2, \ldots, N-M,$$

with $\hat{T}_k$ and $\tilde{T}_k$ defined analogously as in Section 4.5.2. Note the similarity here with the results for the J-M model of Section 4.5.2.

4.5.4 IBM Model

For the IBM model, an analysis of the type discussed in Sections 4.6.2 and 4.6.3 is not appropriate. The reason for this is best explained by examining the distribution of the time to detect the next error, given M prior detections in time t (this is equivalent to S_{M+1}). The event $\{S_{M+1} \geq h\}$ $(h > 0)$ is equivalent to the event {no. of errors in the interval t to t+h is zero}. Since the processes N(t) has independent increments,

$$P\{S_{M+1} \geq h\} = e^{-(m(t+h) - m(t))}$$

so that

$$\lim_{h \to \infty} P\{S_{M+1} \geq h\} = e^{-(a-m(t))}$$

and hence, $S_{M+1} = \infty$ with probability $e^{-(a-m(t))}$ and therefore has infinite expectation. This probability that $S_{M+1} = \infty$ can be made arbitrarily small by taking t large (i.e. by observing/removing errors for a longer time). However, it is clear that an analysis of the type performed earlier will fail.

However, there is a measure of the time t_q required to remove, on the average, a specified fraction q of the expected number of errors ever observable, namely a. To

derive this measure, suppose that the specified fraction is q. On the average, then, the number of errors observable after time t is $E(N(\infty) - N(t)) = a - a(1-e^{-bt}) = ae^{-bt}$. Thus, it is desired that

$$(1-q) = \frac{ae^{-bt_q}}{a} = e^{-bt_q}, \quad 0 < q < 1,$$

or $\quad t_q = -\ln(1-q)/b$. Since

b is unknown, estimates of t_q are $\hat{t}_q$ and $\tilde{t}_q$ where b is replaced by $\hat{b}$ and $\tilde{b}$, respectively.

Section 5.0

MODEL VALIDATION AND DATA ANALYSIS

5.1 PARAMETER ESTIMATES

5.1.1 Model Comparisons

The parameter estimates for data sets 1–16 and for the various model types are given in Tables 5.1.1 – 5.1.16. The notations and abbreviations used in the subject tables can be found in the Glossary given in Section 8.0.

The asterisks used in the tables to indicate the absence of an estimate require some further explanation. A single asterisk indicates that after reasonable diligence with varying starting points we could not obtain convergence of the numerical method to a single parameter point. The cause of this lack of convergence is, in addition to the absence of a rational method of selecting starting points, probably the failure of the data to satisfy an assumption of the model - the expected number of errors per unit time is a decreasing function of time. The reason we say this is that the IBM model never had a convergence problem and it did not assume a decreasing mean error rate; only that the cumulated failures in time is an increasing function of time. <u>All</u> data satisfy that assumption.

The double asterisk indicates a situation in which the numerical method converged to a point but the number was absurd. For example $\hat{N}$ might have been 100 when the number of errors already observed was 200. All of the models suffered from this problem although it was not quite as prevalent as the single asterisk. In this connection the large number of negative $\tilde{\alpha}$ s and $\hat{\alpha}$ s occurring for the GPM should be mentioned. This was not given a double asterisk even though a negative α means the expected number of errors <u>decreases</u> as the interval width <u>increases</u>. This may sound absurd but it violates no explicit assumption. Also it is interesting to note that the negative values are often, indeed most often, not far from zero and the problem could have been caused by the point previously raised - the observed mean error rate per unit time was not decreasing. In the few cases in which a positive $\tilde{\alpha}$ or $\hat{\alpha}$ was obtained the number was nearer zero than one (J-M model) or two (S-W model) lending credence to the belief that, at least for the observed data, the error occurrence rate did not depend on the interval width or at least that dependence was obscured by other more dominant factors.

TABLE 5.1.1. DATA SET 1 – PARAMETER ESTIMATES

Data Set / Model		1-fd	1-fdG	1-fx	1-fxG	1-ff	1-ffG
GPM	$\tilde{N}$				7483		
	$\tilde{\alpha}$	*	*	*	-.3543	*	*
	$\tilde{\phi}$				.0065		
J-M	$\tilde{N}$	**	**	13062	17059	**	**
	$\tilde{\phi}$			$7.03\text{x}10^{-5}$	$5.56\text{x}10^{-6}$		
S-W	$\tilde{N}$	**	*	2491	3190	**	**
	$\tilde{\phi}$			$1.83\text{x}10^{-5}$	$5.98\text{x}10^{-6}$		
GPM	$\hat{N}$				9389		
	$\hat{\alpha}$	*	*	*	-.3654	*	*
	$\hat{\phi}$				.0051		
J-M	$\hat{N}$	*	*	*	*	*	*
	$\hat{\phi}$						
S-W	$\hat{N}$	*	*	*	*	*	*
	$\hat{\phi}$						
Binomial	$\tilde{N}$	*	*	*	13631	***	***
I	$\tilde{c}$				14126		
Binomial	$\tilde{N}$	*	**	**	**	***	***
II	$\tilde{a}$						
IBM	$\tilde{a}$	$1.53\text{x}10^{10}$	$1.58\text{x}10^{10}$	$1.19\text{x}10^{8}$	$1.27\text{x}10^{8}$	***	***
	$\tilde{b}$	$2.15\text{x}10^{-7}$	$2.08\text{x}10^{-7}$	$2.10\text{x}10^{-7}$	$2.12\text{x}10^{-7}$		
IBM	$\hat{a}$	**	$2.30\text{x}10^{11}$	$1.05\text{x}10^{10}$	$7.27\text{x}10^{10}$	***	***
	$\hat{b}$		$4.36\text{x}10^{-10}$	$8.43\text{x}10^{-9}$	$1.22\text{x}10^{-9}$		
No. of Errors		2035	2035	2035	2035	2035	2035
No. of Time Intervals		354	142	235	80	354	142

TABLE 5.1.2. DATA SET 2 – PARAMETER ESTIMATES

Data Set Model		2-fx	2-fxG	+	+	+	+
GPM	$\tilde{N}$		20447				
	$\tilde{\alpha}$	*	-.0742				
	$\tilde{\phi}$		.0012				
J-M	$\tilde{N}$	7916	8004				
	$\tilde{\phi}$	.0020	.0018				
S-W	$\tilde{N}$						
	$\tilde{\phi}$	**	**				
GPM	$\hat{N}$	8065	20491				
	$\hat{\alpha}$	-.1320	-.0755				
	$\hat{\phi}$	.0022	.0012				
J-M	$\hat{N}$	8174	8170				
	$\hat{\phi}$	.0027	.0027				
S-W	$\hat{N}$						
	$\hat{\phi}$	*	**				
Binomial	$\tilde{N}$	7929	8025				
I	$\tilde{c}$	498	553				
Binomial	$\tilde{N}$	7923	8014				
II	$\tilde{a}$	.0020	.0018				
IBM	$\tilde{a}$	5459	4997				
	$\tilde{b}$	.0815	.0837				
IBM	$\hat{a}$	8182	8181				
	$\hat{b}$	.0803	.0804				
No. of Errors		7819	7819				
No. of Time Intervals		834	430				

\+ **No find data available**

TABLE 5.1.3. DATA SET 3 – PARAMETER ESTIMATES

Data Set / Model		3-fd	3-fdG	3-fx	3-fxG	3-ff	3-ffG
GPM	$\tilde{N}$, $\tilde{\alpha}$, $\tilde{\phi}$	*	*	*	*	*	**
J-M	$\tilde{N}$, $\tilde{\phi}$	**	**	**	**	**	**
S-W	$\tilde{N}$, $\tilde{\phi}$	**	**	**	**	**	**
GPM	$\hat{N}$, $\hat{\alpha}$, $\hat{\phi}$	*	*	*	*	*	*
J-M	$\hat{N}$, $\hat{\phi}$	*	*	*	*	*	*
S-W	$\hat{N}$, $\hat{\phi}$	*	*	*	*	*	*
Binomial I	$\tilde{N}$, $\tilde{c}$	*	*	*	*	***	***
Binomial II	$\tilde{N}$, $\tilde{a}$	**	*	*	*	***	***
IBM	$\tilde{a}$	7.51×10^{7}	9.50×10^{7}	9.42×10^{7}	1.03×10^{8}	***	***
	$\tilde{b}$	2.11×10^{-7}	2.08×10^{-7}	2.08×10^{-7}	2.12×10^{-7}		
IBM	$\hat{a}$, $\hat{b}$	**	**	**	**	***	***
No. of Errors		366	366	366	366	366	366
No. of Time Intervals		94	27	82	28	94	27

TABLE 5.1.4. DATA SET 4 – PARAMETER ESTIMATES

Data Set Model		4-fd	4-fdG	4-fx	4-fxG	4-ff	4-ffG
GPM	$\tilde{N}$		17208			13398	15672
	$\tilde{\alpha}$	*	-.0199	*	*	-.1233	-.0138
	$\tilde{\phi}$		.0023			.0023	.0025
J-M	$\tilde{N}$	4297	4246	4141	4181	4261	4232
	$\tilde{\phi}$	.0094	.0078	.0073	.0063	.0092	.0076
S-W	$\tilde{N}$ $\tilde{\phi}$	**	**	**	**	**	*
GPM	$\hat{N}$		14928	34956		11010	13621
	$\hat{\alpha}$	*	-.0389	-.4831	*	-.1996	-.0336
	$\hat{\phi}$		.0027	9.29×10^{-4}		.0029	.0029
J-M	$\hat{N}$	4924	4929	**	**	4453	4453
	$\hat{\phi}$	.0087	.0087			.0100	.0101
S-W	$\hat{N}$ $\hat{\phi}$	*	**	*	*	**	*
Binomial I	$\tilde{N}$	4346	4308	4212	4223	***	***
	$\tilde{c}$	107	128	136	158		
Binomial II	$\tilde{N}$	4322	4277	4198	4204	***	***
	$\tilde{a}$	.0094	.0078	.0073	.0063		
IBM	$\tilde{a}$	3073	2130	2159	1730	***	***
	$\tilde{b}$	.4087	.4670	.4522	.4660		
IBM	$\hat{a}$	4967	4965	4229	4230	***	***
	$\hat{b}$	.2587	.2591	.3363	.3361		
No. of Errors		4176	4176	4176	4176	4176	4176
No. of Time Intervals		161	124	149	121	161	124

TABLE 5.1.5. DATA SET 5 – PARAMETER ESTIMATES

Data Set Model		5-fd	5-fdG	5-fx	5-fxG	5-ff	5-ffG
GPM	$\tilde{N}$	881	1752	1084	2929	706	1673
	$\tilde{\alpha}$	-.6625	-.1184	-.5746	-.0725	-.6600	-.1062
	$\tilde{\phi}$	.0075	.0101	.0055	.0051	.0096	.0102
J-M	$\tilde{N}$	**	**	501	650	**	**
	$\tilde{\phi}$			.0008	6.39×10^{-4}		
S-W	$\tilde{N}$	**	**	458	495	**	**
	$\tilde{\phi}$			1.57×10^{-5}	1.26×10^{-5}		
GPM	$\hat{N}$	735	1638	1590	3011	639	1592
	$\hat{\alpha}$	-.6528	-.1170	-.4381	-.0711	-.6379	-.1043
	$\hat{\phi}$	.0098	.0109	.0034	.0050	.0110	.0107
J-M	$\hat{N}$	*	*	507	508	*	*
	$\hat{\phi}$			.0053	.0052		
S-W	$\hat{N}$	*	*	*	*	*	*
	$\hat{\phi}$						
Binomial	$\tilde{N}$	*	*	505	651	***	***
I	$\tilde{c}$			1202	1456		
Binomial	$\tilde{N}$	*	**	503	650	***	***
II	$\tilde{a}$			.0008	.0007		
IBM	$\tilde{a}$	6.44×10^{7}	8.77×10^{7}	74.64	127.6	***	***
	$\tilde{b}$	2.12×10^{-7}	2.14×10^{-7}	.1758	.1125		
IBM	$\hat{a}$	**	**	513	515	***	***
	$\hat{b}$			.1564	.1543		
No. of Errors		450	450	450	450	450	450
No. of Time Intervals		109	34	118	36	109	34

TABLE 5.1.6. DATA SET 6 – PARAMETER ESTIMATES

Data Set Model		6-fd	6-fdG	X	X	X	X
GPM	$\tilde{N}$						
	$\tilde{\alpha}$	**	**				
	$\tilde{\phi}$						
J-M	$\tilde{N}$	919	1324				
	$\tilde{\phi}$	6.4×10^{-4}	6.99×10^{-4}				
S-W	$\tilde{N}$	636	547				
	$\tilde{\phi}$	5.4×10^{-5}	7.62×10^{-5}				
GPM	$\hat{N}$						
	$\hat{\alpha}$	*	*				
	$\hat{\phi}$						
J-M	$\hat{N}$						
	$\hat{\phi}$	**	**				
S-W	$\hat{N}$	594	875				
	$\hat{\phi}$	4.8×10^{-4}	7.62×10^{-5}				
Binomial	$\tilde{N}$	873	1198				
I	$\tilde{c}$	1440	1239				
Binomial	$\tilde{N}$	895	1253				
II	$\tilde{a}$	.0007	.0008				
IBM	$\tilde{a}$	458	1218				
	$\tilde{b}$	.0384	.0225				
IBM	$\hat{a}$	2.75×10^{9}	1.49×10^{5}				
	$\hat{b}$	1.11×10^{-8}	2.04×10^{-4}				
No. of Errors		406	406				
No. of Time Intervals		134	28				

X No fix data available

TABLE 5.1.7. DATA SET 7 – PARAMETER ESTIMATES

Data Set Model		7-fd	7-fdG	7-fx	7-fxG	7-ff	7-ffG
GPM	$\tilde{N}$	4720	6121	6376	7569		4739
	$\tilde{\alpha}$	-.2854	.2150	.3701	.2481	**	.2328
	$\tilde{\phi}$	.0050	.0042	.0042	.0059		.0048
J-M	$\tilde{N}$	4451	4370	5023	10820	**	**
	$\tilde{\phi}$	.0049	.0046	.0018	6.19×10^{-4}		
S-W	$\tilde{N}$	3787	3657	3978	13641	*	**
	$\tilde{\phi}$	.0045	.0023	.0003	1.82×10^{-5}		
GPM	$\hat{N}$	4618	6179	7805	7341		4869
	$\hat{\alpha}$	-.2552	.2123	.3128	.2934	**	.2240
	$\hat{\phi}$	.0052	.0042	.0034	.0057		.0047
J-M	$\hat{N}$	4458	4454	7457	7434	*	**
	$\hat{\phi}$	.0052	.0052	.0014	.0014		
S-W	$\hat{N}$	4186	3719	6306	9278	*	*
	$\hat{\phi}$	.0053	.0045	.0003	1.03×10^{-4}		
Binomial	$\tilde{N}$	4450	4376	4994	9642	***	***
I	$\tilde{c}$	202	216	534	1386		
Binomial	$\tilde{N}$	4451	4373	5008	6055	***	***
II	$\tilde{a}$	.0049	.0046	.0019	.0007		
IBM	$\tilde{a}$	4203	3715	4178	9276	***	***
	$\tilde{b}$	.1528	.1618	.0662	.0209		
IBM	$\hat{a}$	4470	4464	8844	8426	***	***
	$\hat{b}$	.1562	.1567	.0334	.0356		
No. of Errors		3537	3537	3530	3530	3537	3537
No. of Time Intervals		282	189	123	70	282	189

TABLE 5.1.8. DATA SET 8 – PARAMETER ESTIMATES

Data Set Model		8-fd	+	X	X	X	X
GPM	$\tilde{N}$						
	$\tilde{\alpha}$	*					
	$\tilde{\Phi}$						
J-M	$\tilde{N}$	1241					
	$\tilde{\Phi}$	.1395					
S-W	$\tilde{N}$	1241					
	$\tilde{\Phi}$	.1395					
GPM	$\hat{N}$						
	$\hat{\alpha}$	*					
	$\hat{\Phi}$						
J-M	$\hat{N}$	1324					
	$\hat{\Phi}$	.1210					
S-W	$\hat{N}$	1324					
	$\hat{\Phi}$	.1210					
Binomial	$\tilde{N}$	1241					
I	$\tilde{c}$	6					
Binomial	$\tilde{N}$	1241					
II	$\tilde{a}$	.1503					
IBM	$\tilde{a}$	1218					
	$\tilde{b}$	.1558					
IBM	$\hat{a}$	1348					
	$\hat{b}$	.1240					
No. of Errors		1138					
No. of Time Intervals		15					

X Fix data not available

\+ Grouping not necessary

TABLE 5.1.9. DATA SET 9 – PARAMETER ESTIMATES

Data Set Model		9-fd	+	X	X	X	X
GPM	$\tilde{N}$						
	$\tilde{\alpha}$	*					
	$\tilde{\phi}$						
J-M	$\tilde{N}$	1693					
	$\tilde{\phi}$	.1230					
S-W	$\tilde{N}$	1693					
	$\tilde{\phi}$	.1230					
GPM	$\hat{N}$						
	$\hat{\alpha}$	*					
	$\hat{\phi}$						
J-M	$\hat{N}$	1798					
	$\hat{\phi}$	.1088					
S-W	$\hat{N}$	1798					
	$\hat{\phi}$	.1088					
Binomial	$\tilde{N}$	1693					
I	$\tilde{c}$	7					
Binomial	$\tilde{N}$	1693					
II	$\tilde{a}$	.1313					
IBM	$\tilde{a}$	1689					
	$\tilde{b}$	.1315					
IBM	$\hat{a}$	1828					
	$\hat{b}$	.1112					
No. of Errors		1483					
No. of Time Intervals		15					

X Fix data not available

+ **Grouping not necessary**

TABLE 5.1.10. DATA SET 10 – PARAMETER ESTIMATES

Data Set Model		10-fd	+	X	X	X	X
GPM	$\tilde{N}$						
	$\tilde{\alpha}$	*					
	$\tilde{\phi}$						
J-M	$\tilde{N}$	3492					
	$\tilde{\phi}$	.0909					
S-W	$\tilde{N}$	3492					
	$\tilde{\phi}$	.0909					
GPM	$\hat{N}$						
	$\hat{\alpha}$	*					
	$\hat{\phi}$						
J-M	$\hat{N}$	3798					
	$\hat{\phi}$	.0787					
S-W	$\hat{N}$	3798					
	$\hat{\phi}$	.0787					
Binomial	$\tilde{N}$	3492					
I	$\tilde{c}$	10					
Binomial	$\tilde{N}$	3492					
II	$\tilde{a}$	.0953					
IBM	$\tilde{a}$	3535					
	$\tilde{b}$	.0940					
IBM	$\hat{a}$	3958					
	$\hat{b}$	.0768					
No. of Errors		2707					
No. of Time Intervals		15					

X Fix data not available

+ Grouping not necessary

TABLE 5.1.11. DATA SET 11 – PARAMETER ESTIMATES

Data Set / Model		11-fd	+	X	X	X	X
GPM	$\tilde{N}$						
	$\tilde{\alpha}$	*					
	$\tilde{\phi}$						
J-M	$\tilde{N}$	2989					
	$\tilde{\phi}$	.0946					
S-W	$\tilde{N}$	2989					
	$\tilde{\phi}$	.0946					
GPM	$\hat{N}$						
	$\hat{\alpha}$	*					
	$\hat{\phi}$						
J-M	$\hat{N}$	3278					
	$\hat{\phi}$	.0806					
S-W	$\hat{N}$	3278					
	$\hat{\phi}$	.0806					
Binomial	$\tilde{N}$	2989					
I	$\tilde{c}$	10					
Binomial	$\tilde{N}$	2989					
II	$\tilde{a}$	.0993					
IBM	$\tilde{a}$	3033					
	$\tilde{b}$	.0970					
IBM	$\hat{a}$	3446					
	$\hat{b}$	.0771					
No. of Errors		2362					
No. of Time Intervals		15					

X Fix data not available

\+ Grouping not necessary

TABLE 5.1.12. DATA SET 12 – PARAMETER ESTIMATES

Data Set Model		12-fd	12-fdG	X	X	X	X
GPM	$\tilde{N}$		69				
	$\tilde{\alpha}$	*	-.2276				
	$\tilde{\Phi}$		.3837				
J-M	$\tilde{N}$	*	28				
	$\tilde{\Phi}$		.0059				
S-W	$\tilde{N}$	*	**				
	$\tilde{\Phi}$						
GPM	$\hat{N}$		69				
	$\hat{\alpha}$	*	-.2276				
	$\hat{\Phi}$		.3837				
J-M	$\hat{N}$	31	26				
	$\hat{\Phi}$	.0068	.0079				
S-W	$\hat{N}$ $\hat{\Phi}$	*	*				
Binomial I	$\tilde{N}$	*	31				
	$\tilde{c}$		106				
Binomial II	$\tilde{N}$	*	30				
	$\tilde{a}$		.0074				
IBM	$\tilde{a}$ $\tilde{b}$	**	**				
IBM	$\hat{a}$	32	29				
	$\hat{b}$	.0069	.0087				
No. of Errors		26	26				
No. of Time Intervals		26	3				

X Fix data not available

TABLE 5.1.13. DATA SET 13 – PARAMETER ESTIMATES

Data Set Model		13-fd	13-fdG	13-fx	13-fxG	13-ff	13-ffG
GPM	$\tilde{N}$	2778			115405	2895	
	$\tilde{\alpha}$	-.0544	*	*	.0975	-.0457	**
	$\tilde{\phi}$	.0024			2.38×10^{-4}	.0023	
J-M	$\tilde{N}$	**	**	*	**	**	**
	$\tilde{\phi}$						
S-W	$\tilde{N}$	**	**	*	**	**	**
	$\tilde{\phi}$						
GPM	$\hat{N}$	2497	211922		45335	2631	
	$\hat{\alpha}$	-.0149	-.0740	*	.1108	-.0058	**
	$\hat{\phi}$	.0028	7.59×10^{-5}		5.94×10^{-4}	.0026	
J-M	$\hat{N}$	**	1604	*	1912	1804	**
	$\hat{\phi}$		.0035		.0024	.0026	
S-W	$\hat{N}$	*	**	*	**	**	*
	$\hat{\phi}$						
Binomial I	$\tilde{N}$	**	**	*	**	***	***
	$\tilde{c}$						
Binomial II	$\tilde{N}$	**	**	*	**	***	***
	$\tilde{a}$						
IBM	$\tilde{a}$	764	876	937	1005	***	***
	$\tilde{b}$	.1552	.1263	.1084	.1139		
IBM	$\hat{a}$	1612	1611	1947	1938	***	***
	$\hat{b}$	.1036	.1042	.0705	.0719		
No. of Errors		1572	1572	1789	1789	1572	1572
No. of Time Intervals		360	110	101	52	360	110

TABLE 5.1.14. DATA SET 14 – PARAMETER ESTIMATES

Data Set Model		14-fd	14-fdG	14-fx	14-fxG	14-ff	14-ffG
GPM	$\tilde{N}$	1472		541	3972	2000	
	$\tilde{\alpha}$	.1050	*	-.3404	-.2314	.0873	*
	$\tilde{\phi}$	.0036		.1876	.0223	.0025	
J-M	$\tilde{N}$	542	565	609	793	571	567
	$\tilde{\phi}$	.0075	.0067	.0055	.0043	.0058	.0063
S-W	$\tilde{N}$	*	**	*	633	**	**
	$\tilde{\phi}$				2.07×10^{-4}		
GPM	$\hat{N}$	1310		555	2727	1809	
	$\hat{\alpha}$	.1367	*	-.1605	-.1200	.1161	*
	$\hat{\phi}$	.0042		.1392	.0251	.0028	
J-M	$\hat{N}$	634	635	773	763	**	669
	$\hat{\phi}$	.0083	.0082	.0055	.0056		.0071
S-W	$\hat{N}$	*	**	693	593	515	524
	$\hat{\phi}$			.0004	4.36×10^{-4}	.0045	.0016
Binomial I	$\tilde{N}$	553	579	618	763	***	***
	$\tilde{c}$	130	142	167	194		
Binomial II	$\tilde{N}$	548	572	614	772	***	***
	$\tilde{a}$	.0076	.0069	.0058	.0047		
IBM	$\tilde{a}$	354	374	471	656	***	***
	$\tilde{b}$	.3510	.3168	.2138	.1577		
IBM	$\hat{a}$	663	658	996	842	***	***
	$\hat{b}$	.2305	.2340	.1169	.1511		
No. of Errors		503	503	502	502	503	503
No. of Time Intervals		111	36	19	11	111	36

TABLE 5.1.15. DATA SET 15 – PARAMETER ESTIMATES

Data Set Model		15-fd	15-fdG	15A-fd	15A-fdG	15B-fd	15B-fdG	15C-fd	15C-fdG
GPM	$\tilde{N}$	4116	1217	947			30	280	515
	$\tilde{\alpha}$	.2170	-.2544	-.4052	*	*	-.7193	-.7081	-.1766
	$\tilde{\Phi}$	.0007	.0404	.0273			50.05	.2872	.0790
J-M	$\tilde{N}$	2152	2504	*	**	*	**	641	350
	$\tilde{\Phi}$	7.4×10^{-5}	6.25×10^{-5}					9.57×10^{-5}	3.73×10^{-4}
S-W	$\tilde{N}$	867	2280	*	**	129	**	*	293
	$\tilde{\Phi}$	2.64×10^{-6}	4.05×10^{-7}			8.4×10^{-8}			2.11×10^{-6}
GPM	$\hat{N}$	6857	1360	1226			30	342	575
	$\hat{\alpha}$	.2099	-.2080	-.3606	*	*	-.7193	-.3433	.1264
	$\hat{\Phi}$	.0004	.0294	.0176			50.05	.0758	.0562
J-M	$\hat{N}$	3365	3705	*	*	*	4011	306	302
	$\hat{\Phi}$	6.53×10^{-5}	5.89×10^{-5}				2.73×10^{-6}	8.47×10^{-4}	8.48×10^{-4}
S-W	$\hat{N}$	4392	5632	*	**	256	*	327	269
	$\hat{\Phi}$	1.23×10^{-6}	3.98×10^{-7}			1.56×10^{-7}		8.78×10^{-6}	6.56×10^{-6}
Binomial	$\tilde{N}$							567	350
I	$\tilde{c}$	*	*	*	*	*	*	8669	2427
Binomial	$\tilde{N}$								350
II	$\tilde{a}$	**	**	*	*	**	**	**	.0004
IBM	$\tilde{a}$	1933	1907	12	**	7.77×10^{5}	1.48×10^{6}	204	199
	$\tilde{b}$	.0024	.0025	.1294		2.09×10^{-7}	2.14×10^{-7}	.0093	.0204
IBM	$\hat{a}$	7023	5306	1.59×10^{10}	**	**	3.87×10^{7}	314	309
	$\hat{b}$	9.06×10^{-4}	.0012	2.66×10^{-10}			8.48×10^{-9}	.0243	.0254
No. of Errors		447	447	318	318	25	25	261	261
No. of Time Intervals		88	31	55	22	17	3	55	18

TABLE 5.1.16. DATA SET 16 – PARAMETER ESTIMATES

Data Set Model		16-fd	16-fdG	16-fx	16-fxG	16-ff	16-ffG
GPM	$\tilde{N}$	2572	5181	5368	13385		
	$\tilde{\alpha}$	.1433	.0476	-.1691	-.0317	*	**
	$\tilde{\phi}$	.0126	.0093	.0007	9.78×10^{-4}		
J-M	$\tilde{N}$	**	28191	1596	1515	**	**
	$\tilde{\phi}$		2.6×10^{-5}	.0011	.0015		
S-W	$\tilde{N}$	**	**	1366	**	3607	**
	$\tilde{\phi}$			.0001		2.03×10^{-6}	
GPM	$\hat{N}$	1703	3790	4820	13535		
	$\hat{\alpha}$	.1239	.0459	-.1698	-.0296	*	*
	$\hat{\phi}$	.0253	.0134	.0007	9.63×10^{-4}		
J-M	$\hat{N}$	3698	3813	1796	1793	*	**
	$\hat{\phi}$	5.49×10^{-4}	5.29×10^{-4}	.0017	.0017		
S-W	$\hat{N}$	*	*	**	*	4710	13162
	$\hat{\phi}$					8.92×10^{-6}	1.99×10^{-6}
Binomial	$\tilde{N}$	*	6908	1600	1550	***	***
I	$\tilde{c}$		8672	706	678		
Binomial	$\tilde{N}$	*	**	1598	1533	***	***
II	$\tilde{a}$			.0011	.0015		
IBM	$\tilde{a}$	1.22×10^{8}	1.03×10^{8}	859	943	***	***
	$\tilde{b}$	2.14×10^{-7}	2.10×10^{-7}	.0617	.0733		
IBM	$\hat{a}$	13826	13064	1804	1799	***	***
	$\hat{b}$	.0038	.0040	.0498	.0501		
No. of Errors		1333	1333	1332	1332	1333	1333
No. of Time Intervals		42	28	458	113	42	28

The three asterisks indicate a situation in which the model was inappropriate. For example the three asterisks invariably present in the ff and ffG columns indicate that IBM model can only accommodate one date or time, either find or fix, but not both (as ff and ffG require).

The IBM model appears to yield parameter estimates most often. The few cases of double asterisks for the IBM model were due to convergence to negative estimates of parameters (a, b) which, a priori, must be positive. For example, a negative $\hat{a}$ or $\tilde{a}$ (in the face of a positive estimate of b) would imply a negative mean number of occurrences per arbitrary time interval, an absurdity. In short the estimates lie outside the parameter space. There are other, seeming, absurdities about the IBM model. For example in data set 1 an estimate of a on the order of 10^{10} seems high when only 2035 errors have been observed. This is probably an indication of poor fit. Another problem with the IBM model is illustrated in data set 2 in which the LSE of a, namely $\tilde{a}$, is less than the number of errors observed. Now a is the limiting ($t \rightarrow \infty$) value of the mean number of occurrences. The fact that $\tilde{a}$ is on the order of 5000 when the observed number of errors is 7819 is not, *per se*, absurd. But if the mean number of occurrences is 5000 ($t \rightarrow \infty$), to observe 7819 is extremely unlikely. Here again the IBM model is probably just a poor fit. In spite of these problems the IBM model behaves well overall.

Even though the MLEs are not available for the Binomial model the LSEs for the parameter N, the most important parameter, behave very nicely and always seem to be reasonable estimates.

As previously mentioned the J-M and S-W models are special cases ($\alpha = 1$, $\alpha = 2$ respectively) of the GPM model. One of the disturbing results of the data analysis is that, by and large, the estimate of N is somewhat sensitive to which model is used.

One feature that is true of all the models is clear enough: the parameter estimates remain pretty much the same irrespective of whether the data is grouped ($\geq$ ten per interval) or not. To some extent it is even immaterial, to the parameter estimates, whether the data was find or fix.

5.1.2 Comparison of MLEs and LSEs

There are many properties (consistency, unbiasedness, mean square error, variability) with which to compare estimates; often which properties are chosen for comparison is a matter of personal taste or is dictated by the situation.

In many respects a "must" property is consistency (the estimate converges to the parameter estimated in probability as the "sample size" gets large). Both the LSE and MLE are consistent. It is almost universally agreed that a proper measure of comparison of estimators invites some measure of precision. We give, in Tables 5.1.2.1, 5.1.2.2 and 5.1.2.3 comparisons based on the standard deviation (std. dev.) of the estimators. The observed std. dev. is taken from the simulations discussed in Section 4.4. The asymptotic std. dev. was also discussed in Section 4.4.

TABLE 5.1.2.1. COMPARISON OF MLEs AND LSEs FOR THE J-M AND S-W MODELS FOR ESTIMATING N

	J-M		S-W	
	$\hat{N}$	$\tilde{N}$	$\hat{N}$	$\tilde{N}$
Observed Std. Dev.	16.3	17.6	4.99	5.65
Asymptotic Std. Dev.	15.6	17.6	4.94	5.57

TABLE 5.1.2.2. COMPARISON OF MLEs AND LSEs FOR THE J-M AND S-W MODELS FOR ESTIMATING ϕ

	J-M		S-W	
	$\hat{\phi}$	$\tilde{\phi}$	$\hat{\phi}$	$\tilde{\phi}$
Observed Std. Dev.	0.0010	0.0011	0.0003	0.0004
Asymptotic Std. Dev.	0.0010	0.0011	0.0003	0.0004

TABLE 5.1.2.3. COMPARISON OF MLEs AND LSEs FOR THE IBM MODEL FOR ESTIMATING a

	$\hat{a}$	$\tilde{a}$
Observed Std. Dev.	20.4	21.3
Asymptotic Std. Dev.	19.2	20.7

From the first two tables (5.1.2.1 and 5.1.2.2) it is obvious that for the J-M and S-W models the MLE is superior to the LSE for both N and ϕ, because the observed and asymptotic std. dev.'s are always smaller for the MLE. However, speaking relatively the superiority of the MLE over the LSE is not large enough to be overridden by other considerations; for example if convergence cannot be obtained for the MLE the LSE is perfectly satisfactory. The same kind of results are also true for Table 5.1.2.3 which considers the IBM model.

Another point is worth noting from the first two tables: the std. dev.'s are less (and sometimes a good deal less) for the S-W model than for the J-M model. This does not mean the S-W model is superior to the J-M; after all the decision of which model, if either fits, must be made on goodness of fit grounds. The differing std. dev.'s for the two models merely reflects the dependence of the std. dev. on α.

Tables 5.1.2.4 and 5.1.2.5 give a feeling for the closeness of the MLE and LSEs for selected (so that estimates existed) data sets for fxG data. The fxG data was selected because we felt it led to the best estimates. Mostly, the comparable values are, as expected from the std. dev. results previously given, close. However there is another point, raised ever so slightly by Table 5.1.2.4: the MLEs (hats) seemed to experience slightly more convergence (single asterisk) problems than the LSEs (tildes). In the three models (GPM, J-M, S-W) for which they were both used there were, over the 16 data sets and over the data types (fd, fdG, etc.), 86 cases of single asterisk for the MLE and only 39 cases for the LSE. However if the double asterisk situation is included the two methods of estimation are more comparable.

5.1.3 Estimation of N

All of the models have parameters (parameter values) peculiar to themselves. For example the J-M has $\alpha = 1$ and the S-W has $\alpha = 2$, while the IBM model has no α at all. The universal and important parameter is N (a in the IBM model), for it is the unknown number of errors present in the S/W. Indeed having estimated N with, say $\hat{N}$, then the very important quantity, (unknown) number of errors remaining in the S/W, can be estimated as $\hat{N}$ - number of errors already removed.

Table 5.1.3.1 gives, for fxG data, the various estimates of N for the various models. The IBM model was not included because it really does not estimate N since the parameter a is the expected (mean) number of errors as $t \to \infty$. Only the MLEs are shown in the table. Actually the J-M, S-W and Binomial models give fairly close results while the GPM differs a good deal. It remains clear that the choice of a model must be made on previous experience with the model (how well it fits), not in answers one is looking for.

5.2 GOODNESS-OF-FIT DEVELOPMENT

To test the goodness-of-fit for the S/W reliability models the following statistic was employed:

$$\chi^2 = \sum_{i=1}^{k} \frac{(\mathrm{OBS}_i - \mathrm{EXPECTED}_i)^2}{\mathrm{EXPECTED}_i}$$

where

$$\mathrm{OBS}_i = \begin{cases} N_i & \text{for the GPM, J-M, S-W, and binomial models} \\ Z_i & \text{for the IBM model} \end{cases}$$

TABLE 5.1.2.4. LSE AND MLE PARAMETER ESTIMATES FOR POISSON MODELS

DATA SET		$\tilde{N}$	$\hat{N}$	$\tilde{\phi}$	$\hat{\phi}$	$\tilde{\alpha}$	$\hat{\alpha}$
	1-fxG	7483	9389	.0065	.0051	-.3543	-.3654
	2-fxG	20447	20491	.0012	.0012	-.0742	-.0755
	5-fxG	2929	3011	.0051	.0050	-.0725	-.0711
GPM	7-fxG	7569	7341	.0059	.0057	.2481	.2934
	13-fxG	115405	45335	2.38×10^{-4}	5.94×10^{-4}	.0975	.1108
	14-fxG	3972	2727	.0223	.0251	-.2314	-.1200
	16-fxG	13385	13535	9.78×10^{-4}	9.63×10^{-4}	-.0317	-.0296
	1-fxG	17059	*	5.56×10^{-6}	*		
	2-fxG	8004	8170	.0018	.0027		
	4-fxG	4181	*	.0063	*		
J-M	5-fxG	650	508	6.39×10^{-4}	.0052		
	7-fxG	10820	7434	6.19×10^{-4}	.0014		
	13-fxG	*	1912	*	.0024		
	14-fxG	793	763	.0043	.0056		
	16-fxG	1515	1793	.0015	.0017		
	1-fxG	3190	*	5.98×10^{-6}	*		
S-W	5-fxG	495	*	1.26×10^{-5}	*		
	7-fxG	13641	9278	1.82×10^{-5}	1.03×10^{-4}		
	14-fxG	633	593	2.07×10^{-4}	4.36×10^{-4}		

TABLE 5.1.2.5. LSE AND MLE PARAMETER ESTIMATES FOR THE IBM MODEL

DATA SET	$\tilde{a}$	$\hat{a}$	$\tilde{b}$	$\hat{b}$
1-fxG	1.27×10^{8}	7.27×10^{10}	2.12×10^{-7}	1.22×10^{-9}
2-fxG	4997	8181	.0837	.0804
4-fxG	1730	4230	.4660	.3361
5-fxG	128	515	.1125	.1543
7-fxG	9276	8426	.0209	.0356
13-fxG	1005	1938	.1139	.0719
14-fxG	656	842	.1577	.1511
16-fxG	943	1799	.0733	.0501

TABLE 5.1.3.1. COMPARISON OF MODELS WITH RESPECT TO ESTIMATION OF N

DATA SET	GPM	J-M	S-W	Bin I	Bin II
1-fxG	7483	17059	3190	13631	*
2-fxG	20447	8004	*	8025	8014
4-fxG	*	4181	*	4223	4204
5-fxG	2929	650	495	651	650
7-fxG	7569	10820	13641	9642	6055
13-fxG	115405	*	*	*	*
14-fxG	3972	793	633	763	772
16-fxG	13385	1515	*	1550	1533

$$
\text{EXPECTED}_i = \begin{cases} \phi\,(N - M_{i-1})\tau_i^{\alpha} & \text{for the GPM, J-M } (\alpha = 1) \text{ and S-W } (\alpha = 2) \text{ models} \\ p(\tau_i)\,(N - \sum_{j=1}^{i-1} N_i) & \text{for the binomial models } (p(\tau_i) = \tau_i/(\tau_i + c) \text{ or } p(\tau_i) = 1 - e^{-a\tau_i}) \\ a(e^{-bt_{i-1}} - e^{-bt_i}) & \text{for the IBM model.} \end{cases}
$$

Since the parameters N, ϕ, α, a, b, etc. are unknown, they were replaced by their MLEs and by their LSEs. For the GPM, J-M, S-W and IBM models the distribution of χ^2 is approximately a chi-square distribution with k degrees of freedom when the parameters are known. This is true because each term in the sum is the square of a normalized Poisson random variable and the terms are independent. When the mean of a Poisson is 10 or greater, it is approximately normally distributed so that χ^2 is approximately the sum of the squares of k independent, standard normal random variables (this, of course, is one way of defining the chi-square distribution; i.e., the distribution of the sum of squares of independent standard normal random variables). When the parameters are replaced by their MLEs or LSEs, χ^2 has approximately a chi-square distribution with degrees of freedom equal to k minus 1 degree of freedom for each parameter estimated (i.e., k-3 for GPM, k-2 for J-M, S-W and IBM models). For the binomial models, whenever N is large and $p(\tau_i)$ small, the chi-square approximation should be valid with k-2 degrees of freedom when the parameters are replaced by their MLEs or LSEs.

To verify the chi-square approximation to the distribution of χ^2 when parameters are estimated the distribution of χ^2 was simulated using $k = 25$, $\phi = 0.01$, $N = 300$, $\tau_i = 10$, $t_i = 10i$ ($i=1, 2, \ldots, 25$), $a = 300$, $b = 0.01$ for the J-M, S-W, and IBM models, respectively, for both the cases where the parameter values were replaced by their MLEs and LSEs. The simulated distributions are based on 500 independent observations of χ^2 and the mean and standard deviation are printed after the class interval tabulations. The hypothesized means and standard deviations are 23 and 6.782, respectively (the mean of a chi-square with df degrees of freedom is df and the standard deviation is $\sqrt{2\text{df}}$). Due to sampling error, the observed means, standard deviations, and quantities of the distributions will not coincide exactly with the true means, standard deviations, and quantities. However, close agreement is observed. (These results are shown in Figure 5.2.1).

As mentioned in Section 3.2 there is an advantage to utilizing the increments Z_i rather than the cumulative observations $N(t_i)$ in the IBM model when testing goodness of fit. In fact, it will be shown here that if the cumulative observations are used in the χ^2 statistic (assuming a and b known) the distribution is not a chi-square distribution but rather has the distribution corresponding to that of k times a random variable with a chi-square distribution with one degree of freedom.

82755-29

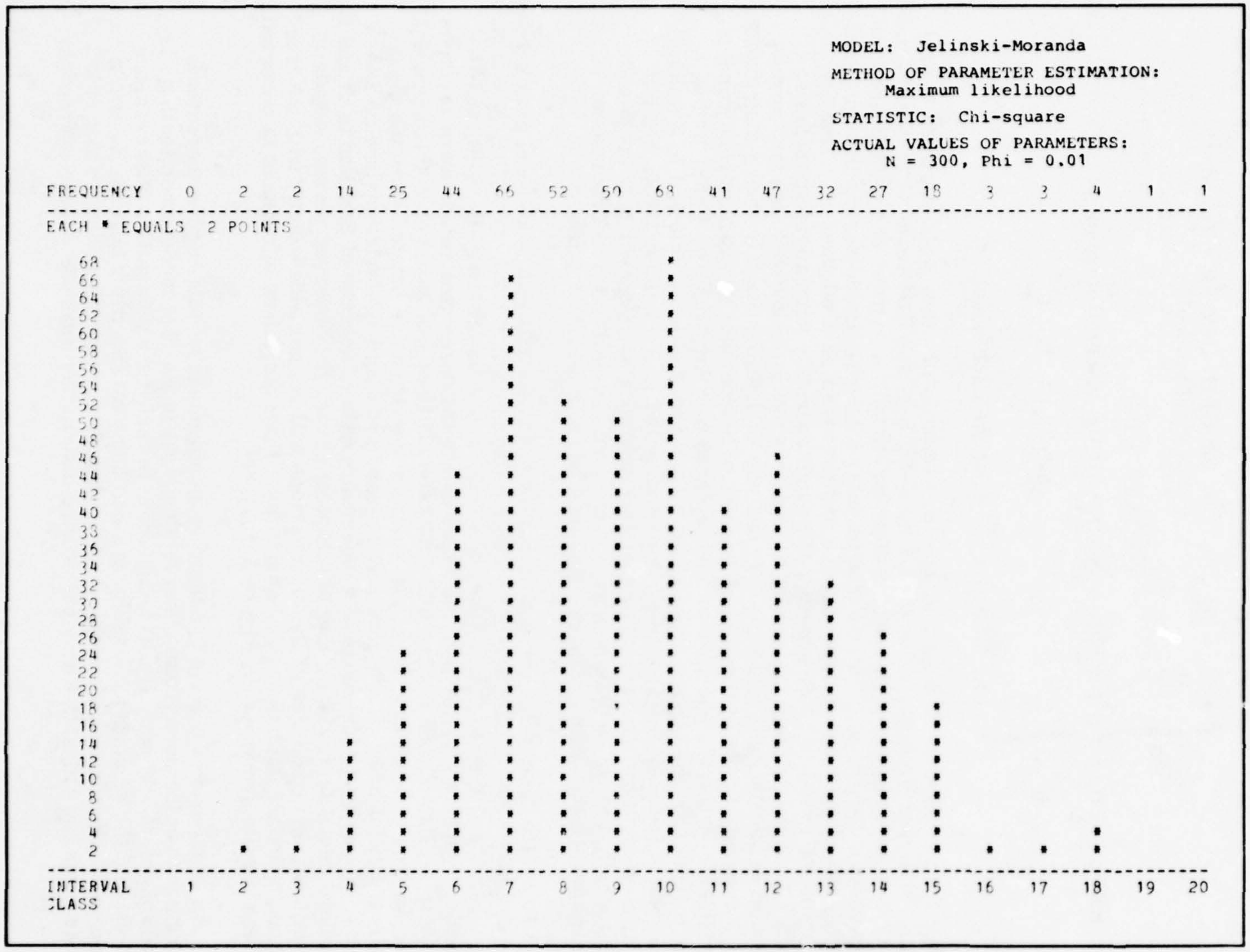

Figure 5.2.1. Simulation of the Chi-Square Statistic for the J-M, S-W, and IBM Models for MLEs and LSEs

82755-30

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.505342E+01
2	0.505342E+01	0.730703E+01
3	0.730703E+01	0.956064E+01
4	0.956064E+01	0.118143E+02
5	0.118143E+02	0.140679E+02
6	0.140679E+02	0.163215E+02
7	0.163215E+02	0.185751E+02
8	0.185751E+02	0.208287E+02
9	0.208287E+02	0.230823E+02
10	0.230823E+02	0.253359E+02
11	0.253359E+02	0.275895E+02
12	0.275895E+02	0.298431E+02
13	0.298431E+02	0.320967E+02
14	0.320967E+02	0.343503E+02
15	0.343503E+02	0.366039E+02
16	0.366039E+02	0.388575E+02
17	0.388575E+02	0.411111E+02
18	0.411111E+02	0.433647E+02
19	0.433647E+02	0.456184E+02
20	GREATER THAN	0.456184E+02

EXPECTED VALUE= 23.0655 STANDARD DEVIATION= 7.0364

Figure 5.2.1. Continued

82755-31

MODEL: Jelinski-Moranda

METHOD OF PARAMETER ESTIMATION:
Least squares

STATISTIC: Chi-square

ACTUAL VALUES OF PARAMETERS:
N = 300, Phi = 0.01

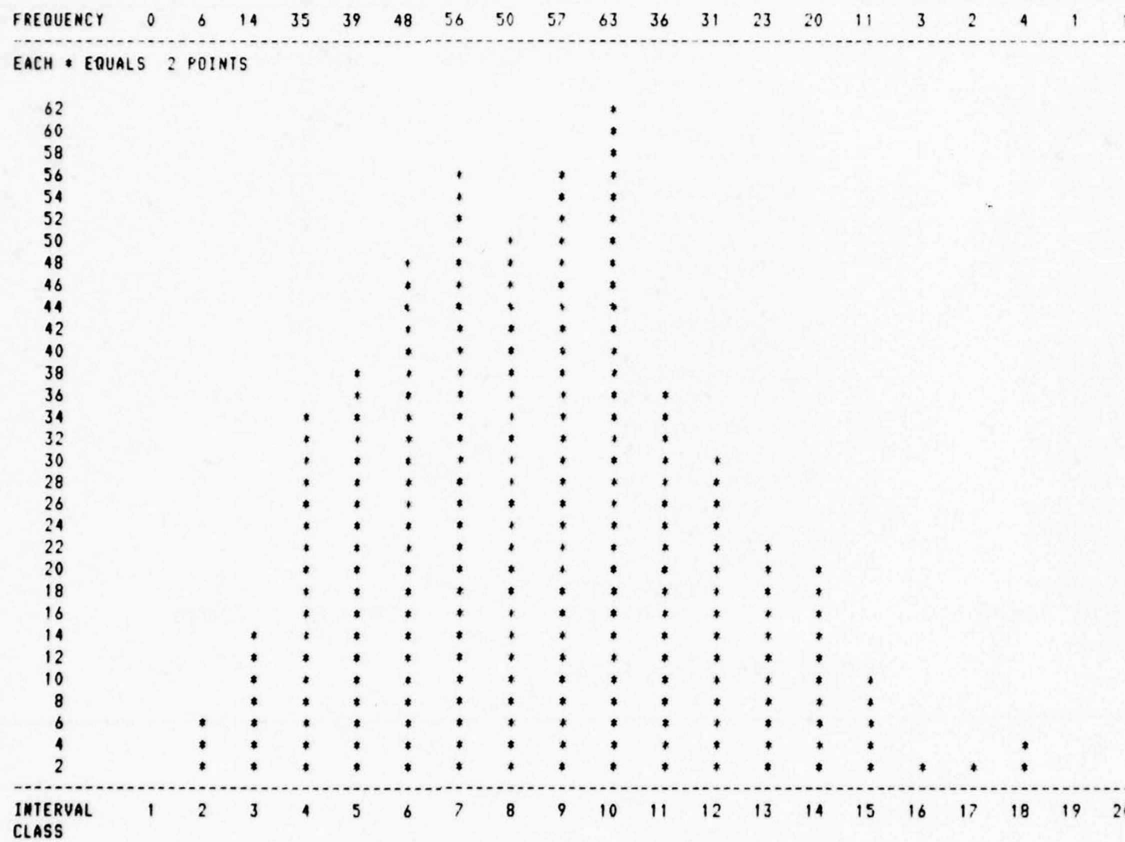

Figure 5.2.1. Continued

82755-32

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.842706E+01
2	0.842706E+01	0.105067E+02
3	0.105067E+02	0.125863E+02
4	0.125863E+02	0.146659E+02
5	0.146659E+02	0.167455E+02
6	0.167455E+02	0.188251E+02
7	0.188251E+02	0.209047E+02
8	0.209047E+02	0.229843E+02
9	0.229843E+02	0.250639E+02
10	0.250639E+02	0.271435E+02
11	0.271435E+02	0.292231E+02
12	0.292231E+02	0.313027E+02
13	0.313027E+02	0.333823E+02
14	0.333823E+02	0.354619E+02
15	0.354619E+02	0.375415E+02
16	0.375415E+02	0.396212E+02
17	0.396212E+02	0.417008E+02
18	0.417008E+02	0.437804E+02
19	0.437804E+02	0.458601E+02
20	GREATER THAN	0.458601E+02

OBSERVED DISTRIBUTION OF A

OBSERVED VALUE OF MEAN = 23.2412
OBSERVED VALUE OF STANDARD DEVIATION = 6.8763

Figure 5.2.1. Continued

82755-33

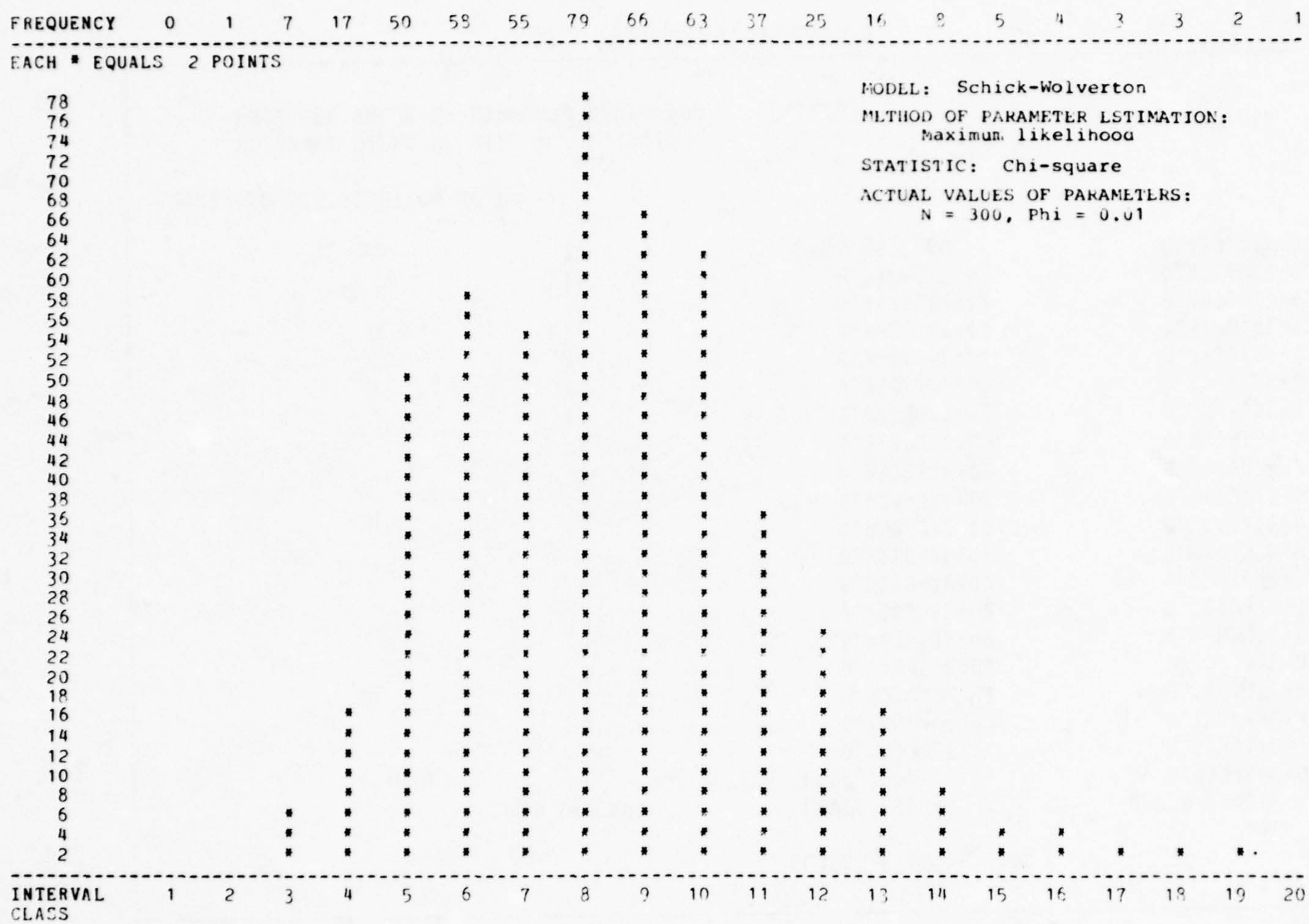

Figure 5.2.1. Continued

82755-34

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.566181E+01
2	0.566181E+01	0.802155E+01
3	0.802155E+01	0.103813E+02
4	0.103813E+02	0.127410E+02
5	0.127410E+02	0.151008E+02
6	0.151008E+02	0.174605E+02
7	0.174605E+02	0.198203E+02
8	0.198203E+02	0.221800E+02
9	0.221800E+02	0.245397E+02
10	0.245397E+02	0.268995E+02
11	0.268995E+02	0.292592E+02
12	0.292592E+02	0.316190E+02
13	0.316190E+02	0.339787E+02
14	0.339787E+02	0.363385E+02
15	0.363385E+02	0.386982E+02
16	0.386982E+02	0.410579E+02
17	0.410579E+02	0.434177E+02
18	0.434177E+02	0.457774E+02
19	0.457774E+02	0.481372E+02
20	GREATER THAN	0.481372E+02

EXPECTED VALUE= 22.2065 STANDARD DEVIATION= 6.8156

Figure 5.2.1. Continued

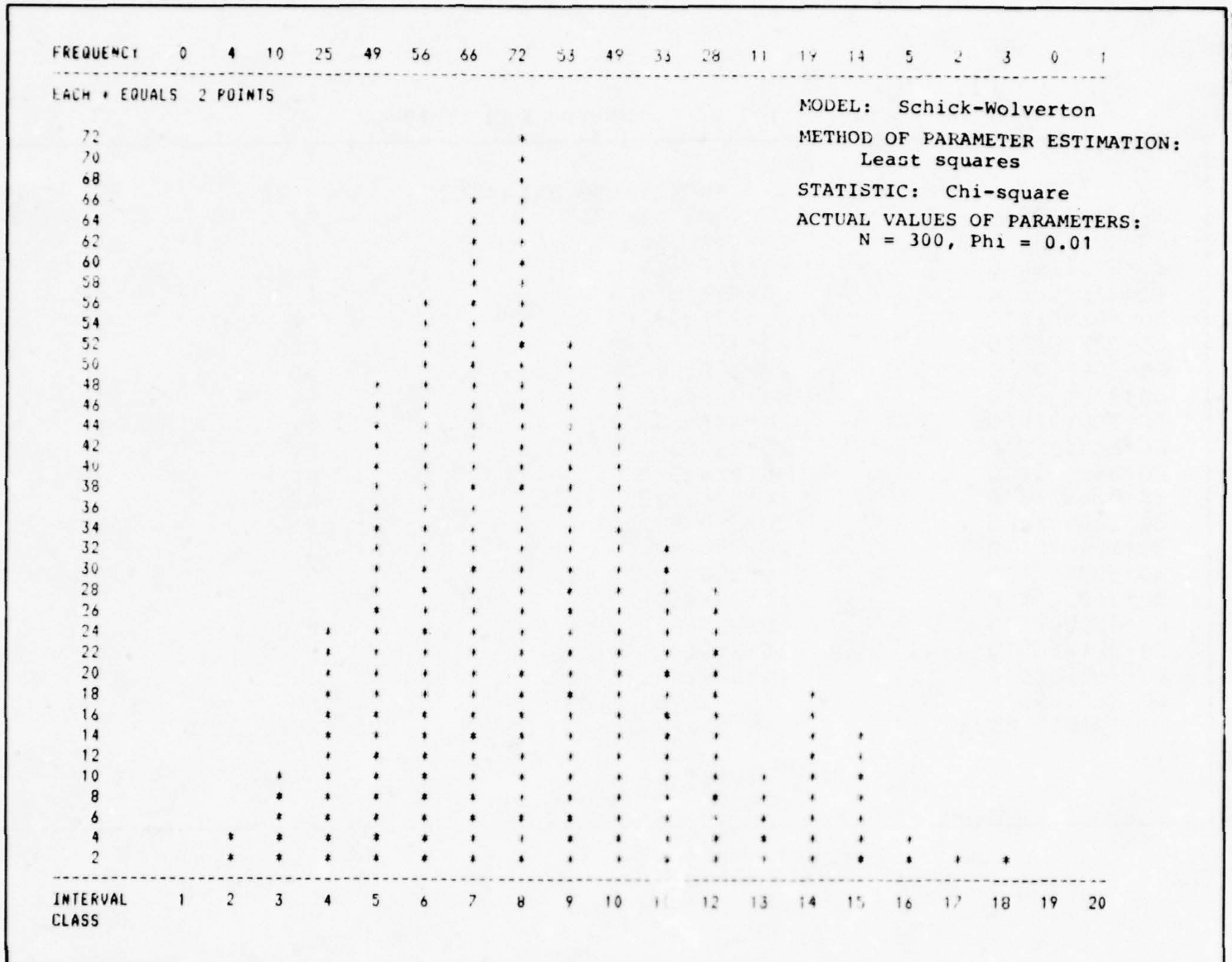

Figure 5.2.1. Continued

82755-36

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.748399E+01
2	0.748399E+01	0.966851E+01
3	0.966851E+01	0.118530E+02
4	0.118530E+02	0.140376E+02
5	0.140376E+02	0.162221E+02
6	0.162221E+02	0.184066E+02
7	0.184066E+02	0.205911E+02
8	0.205911E+02	0.227756E+02
9	0.227756E+02	0.249601E+02
10	0.249601E+02	0.271446E+02
11	0.271446E+02	0.293291E+02
12	0.293291E+02	0.315136E+02
13	0.315136E+02	0.336981E+02
14	0.336981E+02	0.358826E+02
15	0.358826E+02	0.380672E+02
16	0.380672E+02	0.402517E+02
17	0.402517E+02	0.424362E+02
18	0.424362E+02	0.446207E+02
19	0.446207E+02	0.468054E+02
20	GREATER THAN	0.468054E+02

OBSERVED DISTRIBUTION OF A

OBSERVED VALUE OF MEAN = 22.6890
OBSERVED VALUE OF STANDARD DEVIATION = 6.7924

Figure 5.2.1. Continued

82755-37

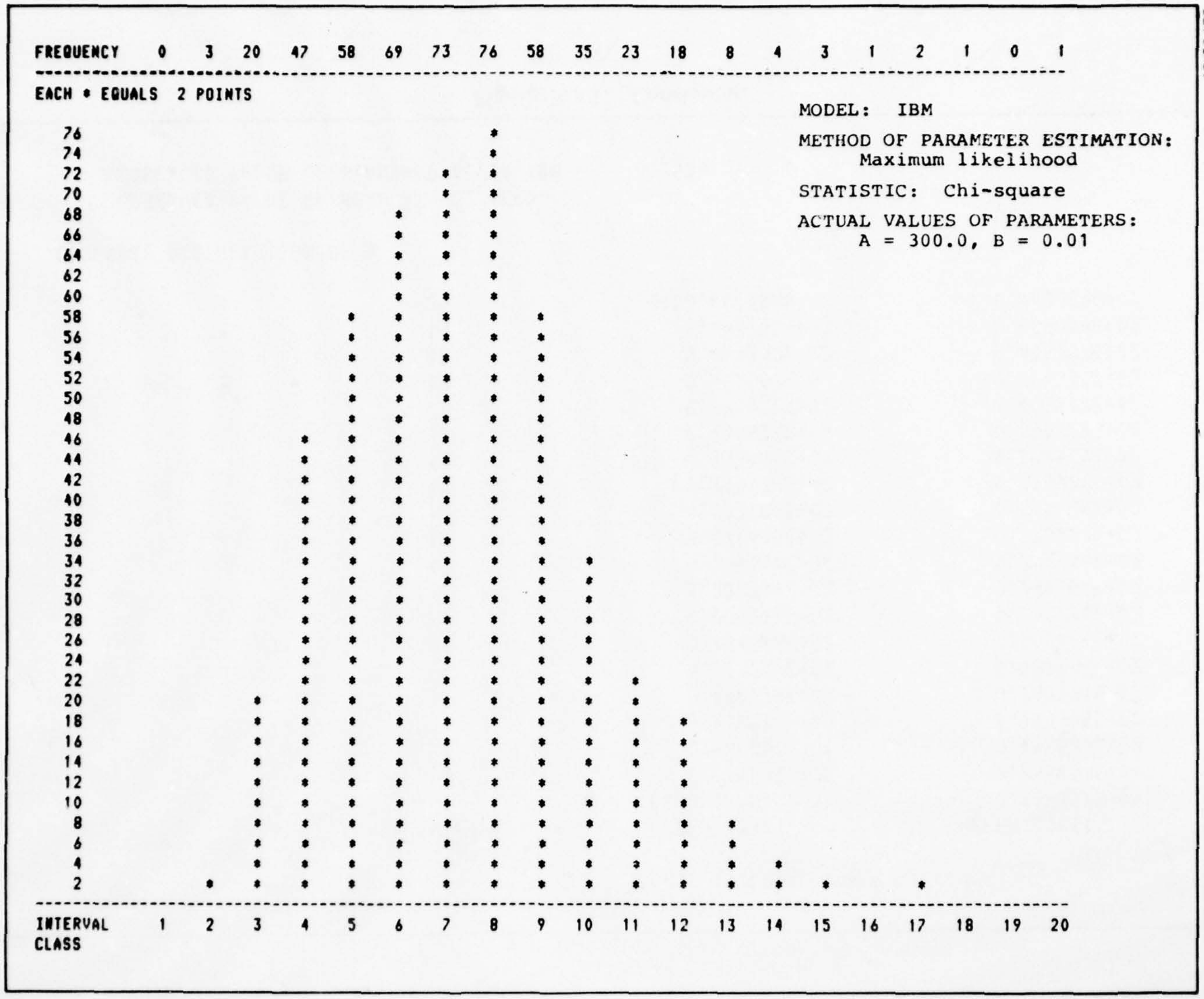

Figure 5.2.1. Continued

82755-38

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.812043E+01
2	0.812043E+01	0.106206E+02
3	0.106206E+02	0.131209E+02
4	0.131209E+02	0.156211E+02
5	0.156211E+02	0.181213E+02
6	0.181213E+02	0.206215E+02
7	0.206215E+02	0.231217E+02
8	0.231217E+02	0.256219E+02
9	0.256219E+02	0.281222E+02
10	0.281222E+02	0.306224E+02
11	0.306224E+02	0.331226E+02
12	0.331226E+02	0.356228E+02
13	0.356228E+02	0.381230E+02
14	0.381230E+02	0.406232E+02
15	0.406232E+02	0.431234E+02
16	0.431234E+02	0.456237E+02
17	0.456237E+02	0.481239E+02
18	0.481239E+02	0.506241E+02
19	0.506241E+02	0.531245E+02
20	GREATER THAN	0.531245E+02

OBSERVED DISTRIBUTION OF CHISQ

OBSERVED VALUE OF MEAN = 22.9307
OBSERVED VALUE OF STANDARD DEVIATION = 6.7573

Figure 5.2.1. Continued

82755-39

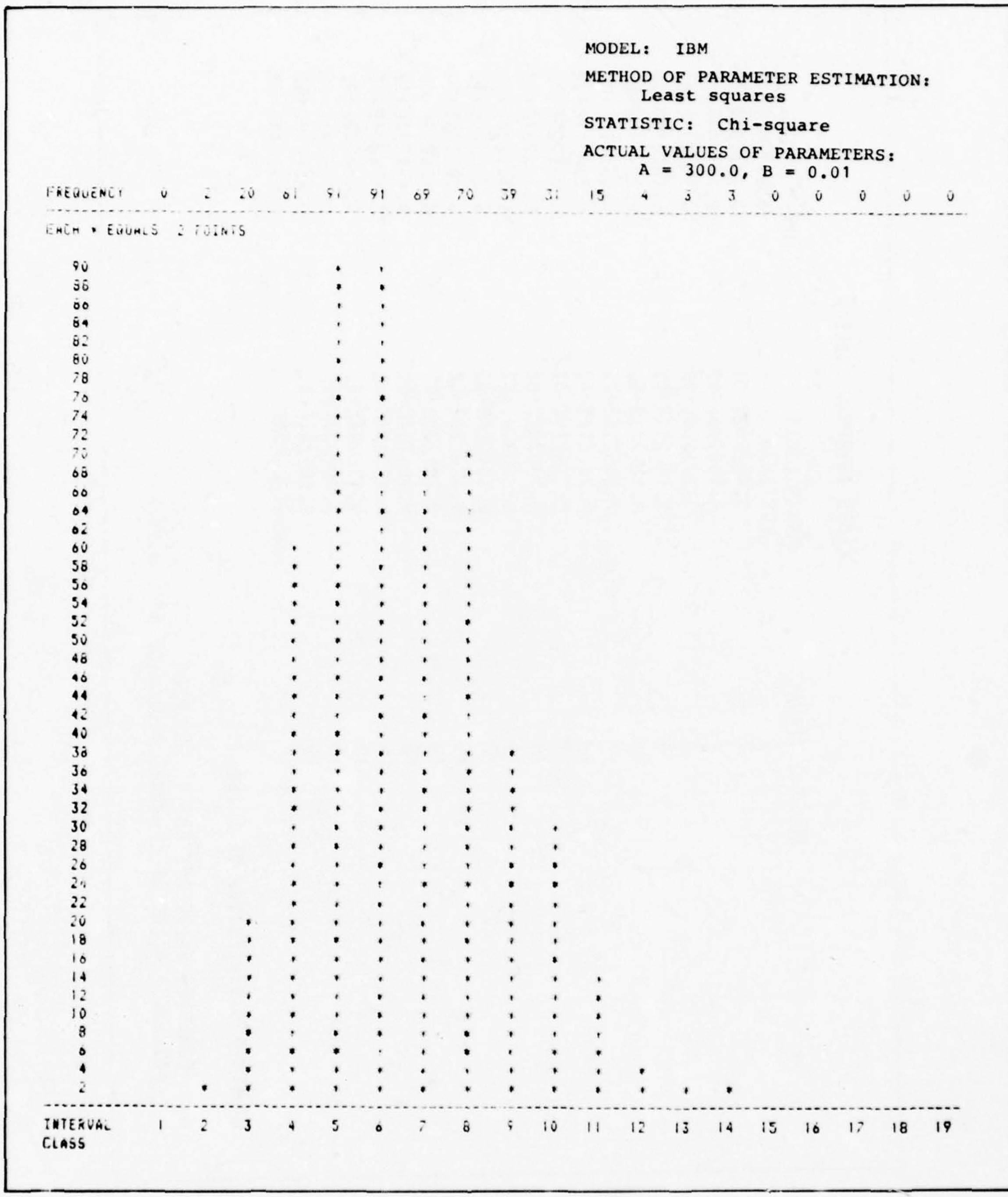

Figure 5.2.1. Continued

82755-40

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.680206E+01
2	0.680206E+01	0.100679E+02
3	0.100679E+02	0.133337E+02
4	0.133337E+02	0.165995E+02
5	0.165995E+02	0.198653E+02
6	0.198653E+02	0.231311E+02
7	0.231311E+02	0.263969E+02
8	0.263969E+02	0.296626E+02
9	0.296626E+02	0.329284E+02
10	0.329284E+02	0.361942E+02
11	0.361942E+02	0.394600E+02
12	0.394600E+02	0.427258E+02
13	0.427258E+02	0.459916E+02
14	0.459916E+02	0.492574E+02
15	0.492574E+02	0.525232E+02
16	0.525232E+02	0.557890E+02
17	0.557890E+02	0.590548E+02
18	0.590548E+02	0.623206E+02
19	0.623206E+02	0.655865E+02
20	GREATER THAN	0.655865E+02

OBSERVED DISTRIBUTION OF CHISQ

OBSERVED VALUE OF MEAN = 23.6015

OBSERVED VALUE OF STANDARD DEVIATION = 7.3361

Figure 5.2.1. Continued

(This can be interpreted as only utilizing one data point and this reflects the built-in stability of cumulative observation models). To show this consider

$$S = \sum_{i=1}^{k} \frac{(N(t_i) - m(t_i))^2}{m(t_i)}$$

where $m(t_i) = a(i-e^{-bt_i})$ (i.e., this is the χ^2 statistic utilizing the cumulative observation rather than the increments Z_i). Consider S^* defined by

$$S^* = \frac{1}{k} S = \sum_{i=1}^{k} \frac{(N(t_i) - m(t_i))^2}{k\, m(t_i)}$$

For $m(t_i) \geq 10$ for each i $N(t_i)$ has an approximate normal distribution with mean $m(t_i)$ and variance $m(t_i)$ so that the vector

$$Y = \begin{pmatrix} \dfrac{N(t_1) - m(t_1)}{\sqrt{k\, m(t_1)}} \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{N(t_k) - m(t_k)}{\sqrt{k\, m(t_k)}} \end{pmatrix}$$

has an approximate k-variate normal distribution with mean equal to the zero vector and covariance matrix (see Section 3.2) equal to

$$\frac{1}{k} \begin{pmatrix} 1 & \cdot & \cdot & \cdot & 1 \\ \cdot & 1 & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ 1 & \cdot & \cdot & \cdot & 1 \end{pmatrix}$$

i.e., $\frac{1}{k}$ times the kxk matrix of ones.

It is well known that (c.f. (I,3) p. 188) if Y has a multivariate normal distribution with mean zero and covariance matrix Σ then a necessary and sufficient condition for Y^TAY (T denotes transpose, A is a kxk matrix) to have a chi-square distribution is that

$$\Sigma A \Sigma A \Sigma = \Sigma A \Sigma$$

in which case the degrees of freedom is equal to the rank of $A\Sigma$. If we take

$$\Sigma = \frac{1}{k}\begin{pmatrix} 1 & \cdot & \cdot & \cdot & 1 \\ \cdot & 1 & & & \cdot \\ \cdot & & \ddots & & \cdot \\ \cdot & & & \cdot & \cdot \\ 1 & \cdot & \cdot & \cdot & 1 \end{pmatrix}$$

$A = I =$ kxk identity matrix, then

$$S^* = Y^TAY = Y^TY$$

and the necessary and sufficient condition is satisfied since

$$\Sigma A \Sigma A \Sigma = \Sigma^3 = \Sigma A \Sigma = \Sigma^2.$$

The rank of $A\,\Sigma = \Sigma$ is clearly 1. Hence, S^* has a chi-square distribution with 1 degree of freedom, i.e. $S = kS^*$ has the distribution described earlier. The probability density function of S^* is

$$f_{S^*}(s^*) = e^{-s^*/2} \Big/ \sqrt{2\pi s^*}, \quad s^* > 0$$

and the density of S is

$$f_S(s) = e^{-s/2k} \Big/ \sqrt{2\pi ks}, \quad s > 0.$$

The mean of S is k and the standard deviation is $\sqrt{2k^2}$. To verify these findings, the distribution of S was simulated for $k = 25$ and is shown in Figure 5.2.2. Note the long tail and asymptotic behavior near zero. The hypothetical mean is $k = 25$ and the variance is $\sqrt{2k^2} = \sqrt{2(25)^2} = 35.355$.

Using the results described in this section the goodness of fit of the S/W reliability models can be tested by computing the χ^2 statistic and comparing its computed value to selected quantities of the chi-square distribution with approppropriate degrees of freedom. If a value of χ^2 is observed larger than say the α quantile ($\alpha = 0.90$, 0.95, etc.) of the appropriate chi-square distribution then the hypothesis that the data actually were generated by the process in question is rejected at the $(1 - \alpha)$ significance level.

MODEL: IBM

STATISTIC: Chi-square

$$\sum \frac{(N(T(I))-M(T(I)))^2}{M(T(I))}$$

```
FREQUENCY    1 3242  869  416  197  131   58   41   17    7    8    8    2    2    0    0    0    0    0
-------------------------------------------------------------------------------------------------------
EACH * EQUALS 65 POINTS

  3185            *
  3120            *
  3055            *
  2990            *
  2925            *
  2860            *
  2795            *
  2730            *
  2665            *
  2600            *
  2535            *
  2470            *
  2405            *
  2340            *
  2275            *
  2210            *
  2145            *
  2080            *
  2015            *
  1950            *
  1885            *
  1820            *
  1755            *
  1690            *
  1625            *
  1560            *
  1495            *
  1430            *
  1365            *
  1300            *
  1235            *
  1170            *
  1105            *
  1040            *
   975            *
   910            *
   845            *    *
   780            *    *
   715            *    *
   650            *    *
   585            *    *
   520            *    *
   455            *    *
   390            *    *    *
   325            *    *    *
   260            *    *    *
   195            *    *    *    *
   130            *    *    *    *    *
    65            *    *    *    *    *
-------------------------------------------------------------------------------------------------------
INTERVAL     1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16   17   18   19
CLASS
```

82755-41

Figure 5.2.2. Simulation of the Chi-Square Statistic for the IBM Model Based on Cumulative Failures

82755-42

CLASS INTERVAL TABLE

CLASS INTERVAL	LOWER LIMIT	UPPER LIMIT
1	LESS THAN	0.487538E+00
2	0.487538E+00	0.219576E+02
3	0.219576E+02	0.434276E+02
4	0.434276E+02	0.648976E+02
5	0.648976E+02	0.863677E+02
6	0.863677E+02	0.107838E+03
7	0.107838E+03	0.129308E+03
8	0.129308E+03	0.150778E+03
9	0.150778E+03	0.172248E+03
10	0.172248E+03	0.193718E+03
11	0.193718E+03	0.215188E+03
12	0.215188E+03	0.236658E+03
13	0.236658E+03	0.258128E+03
14	0.258128E+03	0.279598E+03
15	0.279598E+03	0.301068E+03
16	0.301068E+03	0.322538E+03
17	0.322538E+03	0.344008E+03
18	0.344008E+03	0.365478E+03
19	0.365478E+03	0.386948E+03
20	GREATER THAN	0.386948E+03

EXPECTED VALUE= 25.1208 STANDARD DEVIATION= 31.5878

Figure 5.2.2. Continued

5.3 GOODNESS-OF-FIT RESULTS

It is desired to test the hypothesis that for a given model, the data under consideration were actually generated by a process governed by that model. This hypothesis will be denoted by H_0. The alternative hypothesis will simply be that H_0 is false.

In Section 5.2 it was shown that when H_0 is true, the statistic χ^2 discussed there has a certain chi-square distribution and can thus be used to test H_0. The hypothesis H_0 is rejected if a significantly large value of χ^2 is observed. By "significantly large" it is meant that the observed χ^2 is larger than a preselected β-quantile of the appropriate chi-square distribution ($0 < \beta < 1$). The value $1-\beta$ is called the significance level.

Tables 5.3.1 through 5.3.16 give the χ^2 values for testing goodness of fit. Below each χ^2 value is a normalized χ^2 value defined by

$$(\chi^2 - df)/\sqrt{2\,df}$$

which for large df has a N(0, 1) distribution (N(0, 1) is a normal distribution with mean zero and variance one) where df is the degrees of freedom of the chi-square distribution which χ^2 has when H_0 is true (df = the number of debugging-time intervals less the number of parameters estimated from the data). Whenever df is so large (greater than or equal to 100) that chi-square tables are not available the normalized χ^2 value can be compared to the β-quantile of the N(0, 1) distribution, since when H_0 is true and df large, the normalized χ^2 value is distributed as N(0, 1), approximately.

To illustrate these concepts consider Table 5.3.15 referring to the GPM (LSE) fit to the 15C-fdG data. The χ^2 value is 28 with degrees of freedom df = 15 (see bottom of table for number of time intervals). The .975 quantile of the chi-square distribution with 15 degrees of freedom is 27.5 so that the χ^2 value of 28 is significant at the .025 significance level and H_0 would be rejected at this level. The .99 quantile of the chi-square distribution with 15 degrees of freedom is 30.6 so that the χ^2 value of 28 is not significant at the .01 significance level.

The significance level represents the probability of rejecting H_0 when it is in fact true and thus must be determined subjectively.

On the basis of Tables 5.3.1 through 5.3.16 it is seen that H_0 would be rejected in almost every case at any reasonable significance level. However, some exceptional cases are worth pointing out.

The GPM, J-M and S-W models appeared to be the models which achieved the best fits. Table 5.3.17 summarizes the best fits achieved by these models. The majority of these "good" fits were for the GPM. This is not surprising since the GPM has three parameters and hence extra freedom to fit. The data represented in Table 5.3.17 contain data from a command and control system (5), a shipboard radar (14), a sonar system (15) and a ground fixed radar (16). No other models exhibited wider applicability than the GPM, J-M, and S-W models.

TABLE 5.3.1. CHI-SQUARE VALUES FOR DATA SET 1

Data Set Model		1-fd	1-fdG	1-fx	1-fxG	1-ff	1-ffG
GPM	(LSE)	*	*	*	1356 103	*	*
J-M	(LSE)	*	*	86622 4002	52441 4192	*	*
S-W	(LSE)	*	*	$2.86x10^6$ 132324	$2.95x10^6$ 236132	*	*
GPM	(MLE)	*	*	*	1360 103	*	*
J-M	(MLE)	*	*	*	*	*	*
S-W	(MLE)	*	*	*	*	*	*
Binomial I	(LSE)	*	*	*	52390 4188	***	***
Binomial II	(LSE)	*	*	*	*	***	***
IBM	(LSE)	20235 748	12189 718	95470 4402	53967 4287	***	***
IBM	(MLE)	*	3082 176	25873 1188	15351 1223	***	***
No. of Time Intervals		354	142	235	80	354	142

TABLE 5.3.2. CHI-SQUARE VALUES FOR DATA SET 2

Data Set Model		2-fx	2-fxG
GPM	(LSE)	*	1392 33
J-M	(LSE)	15500 360	10698 351
S-W	(LSE)	*	*
GPM	(MLE)	13216 304	1392 33
J-M	(MLE)	7524 164	5044 158
S-W	(MLE)	*	*
Binomial I	(LSE)	15040 348	10172 333
Binomial II	(LSE)	15402 357	10424 342
IBM	(LSE)	13041 299	10741 352
IBM	(MLE)	7779 170	5314 167
No. Time Intervals		834	430

TABLE 5.3.3. CHI-SQUARE VALUES FOR DATA SET 3

Data Set Model		3-fd	3-fdG	3-fx	3-fxG	3-ff	3-ffG
GPM	(LSE)	*	*	*	*	*	*
J-M	(LSE)	*	*	*	*	*	*
S-W	(LSE)	*	*	*	*	*	*
GPM	(MLE)	*	*	*	*	*	*
J-M	(MLE)	*	*	*	*	*	*
S-W	(MLE)	*	*	*	*	*	*
Binomial I	(LSE)	*	*	*	*	***	***
Binomial II	(LSE)	*	*	*	*	***	***
IBM	(LSE)	5792 418	2246 308	2987 228	1851 248	***	***
IBM	(MLE)	*	*	*	*	***	***
No. Time Intervals		94	27	82	28	94	27

TABLE 5.3.4. CHI-SQUARE VALUES FOR DATA SET 4

Data Set Model		4-fd	4-fdG	4-fx	4-fxG	4-ff	4-ffG
GPM	(LSE)	*	1445 85	*	*	2827 150	1440 85
J-M	(LSE)	8185 450	11112 704	30517 1771	23943 1544	5453 297	6540 411
S-W	(LSE)	*	*	*	*	*	*
GPM	(MLE)	*	1448 85	3191 178	*	2854 152	1442 85
J-M	(MLE)	4205 227	3691 229	*	*	4047 218	3532 218
S-W	(MLE)	*	*	*	*	*	*
Binomial I	(LSE)	7446 409	9339 590	27453 1592	21171 1365	***	***
Binomial II	(LSE)	7778 427	10101 639	28740 1668	22299 1438	***	***
IBM	(LSE)	7674 420	12408 783	27557 1593	27753 1784	***	***
IBM	(MLE)	4045 218	3494 216	10231 588	7654 488	***	***
No. Time Intervals		161	124	149	121	161	124

TABLE 5.3.5. CHI-SQUARE VALUES FOR DATA SET 5

Data Set Model		5-fd	5-fdG	5-fx	5-fxG	5-ff	5-ffG
GPM	(LSE)	369 18	47 2	389 18	26 -0.88	343 16	47 2
J-M	(LSE)	*	*	12872 837	7598 917	*	*
S-W	(LSE)	*	*	844847 55459	400039 48508	*	*
GPM	(MLE)	358 17	47 2	372 17	26 -.0.88	336 16	47 2
J-M	(MLE)	*	*	1574 96	1139 134	*	*
S-W	(MLE)	*	*	*	*	*	*
Binomial I	(LSE)	*	*	12274 798	7008 846	***	***
Binomial II	(LSE)	*	*	12568 817	*	***	***
IBM	(LSE)	7498 503	4065 496	13986 907	7555 899	***	***
IBM	(MLE)	*	*	1786 110	1340 158	***	***
No. Time Intervals		109	34	118	36	109	34

TABLE 5.3.6. CHI-SQUARE VALUES FOR DATA SET 6

Data Set Model		6-fd	6-fdG
GPM	(LSE)	*	*
J-M	(LSE)	5484 329	1129 153
S-W	(LSE)	104828 6444	37502 5197
GPM	(MLE)	*	*
J-M	(MLE)	*	*
S-W	(MLE)	12863 784	17297 2395
Binomial I	(LSE)	5455 328	1109 150
Binomial II	(LSE)	5470 329	1120 152
IBM	(LSE)	5340 319	945 125
IBM	(MLE)	2110 122	676 90
No. Time Intervals		134	28

TABLE 5.3.7. CHI-SQUARE VALUES FOR DATA SET 7

Data Set Model		7-fd	7-fdG	7-fx	7-fxG	7-ff	7-ffG
GPM	(LSE)	1124 36	376 10	11260 719	3071 260	*	393 11
J-M	(LSE)	1404 48	1065 45	38806 2487	6075 515	*	*
S-W	(LSE)	3737 146	8662 438	536298 34467	76599 6562	*	*
GPM	(MLE)	1124 36	376 10	9952 635	3055 258	*	393 11
J-M	(MLE)	1301 43	806 32	22821 1459	4157 351	*	*
S-W	(MLE)	1823 65	3081 150	136024 8736	19663 1680	*	*
Binomial I	(LSE)	1399 47	1042 44	38460 2464	6030 511	***	***
Binomial II	(LSE)	1402 47	1053 45	38634 2476	6054 513	***	***
IBM	(LSE)	1479 51	1115 48	33210 2118	6234 525	***	***
IBM	(MLE)	1362 46	851 34	19966 1276	4280 361	***	***
No. Time Intervals		282	189	123	70	282	189

TABLE 5.3.8. CHI-SQUARE VALUES FOR DATA SET 8

Data Set Model		8-fd
GPM	(LSE)	*
J-M	(LSE)	186 34
S-W	(LSE)	186 34
GPM	(MLE)	*
J-M	(MLE)	165 30
S-W	(MLE)	165 30
Binomial I	(LSE)	186 34
Binomial II	(LSE)	186 34
IBM	(LSE)	225 40
IBM	(MLE)	182 33
No. Time Intervals		15

TABLE 5.3.9. CHI-SQUARE VALUES FOR DATA SET 9

Data Set Model		9-fd
GPM	(LSE)	*
J-M	(LSE)	167 30
S-W	(LSE)	167 30
GPM	(MLE)	*
J-M	(MLE)	156 28
S-W	(MLE)	156 28
Binomial I	(LSE)	*
Binomial II	(LSE)	167 30
IBM	(LSE)	161 28
IBM	(MLE)	140 25
No. Time Intervals		15

TABLE 5.3.10. CHI-SQUARE VALUES FOR DATA SET 10

Data Set Model		10-fd
GPM	(LSE)	*
J-M	(LSE)	503 96
S-W	(LSE)	503 96
GPM	(MLE)	*
J-M	(MLE)	480 92
S-W	(MLE)	480 92
Binomial I	(LSE)	503 96
Binomial II	(LSE)	503 96
IBM	(LSE)	564 104
IBM	(MLE)	519 99
No. Time Intervals		15

TABLE 5.3.11. CHI-SQUARE VALUES FOR DATA SET 11

Data Set Model		11-fd
GPM	(LSE)	*
J-M	(LSE)	520 99
S-W	(LSE)	520 99
GPM	(MLE)	*
J-M	(MLE)	491 94
S-W	(MLE)	491 94
Binomial	(LSE)	520 99
Binomial II	(LSE)	520 99
IBM	(LSE)	585 108
IBM	(MLE)	530 101
No. Time Intervals		15

TABLE 5.3.12. CHI-SQUARE VALUES FOR DATA SET 12

Data Set Model		12-fd	12-fdG
GPM	(LSE)	*	*
J-M	(LSE)	*	11 7
S-W	(LSE)	*	*
GPM	(MLE)	*	*
J-M	(MLE)	72 7	8 5
S-W	(MLE)	*	*
Binomial I	(LSE)	*	5 3
Binomial II	(LSE)	*	7 4
IBM	(LSE)	*	*
IBM	(MLE)	72 7	10 6
No. Time Intervals		26	3

TABLE 5.3.13. CHI-SQUARE VALUES FOR DATA SET 13

Data Set Model		13-fd	13-fdG	13-fx	13-fxG	13-ff	13-ffG
GPM	(LSE)	1992 61	*	*	487 44	1975 61	*
J-M	(LSE)	*	*	*	*	*	*
S-W	(LSE)	*	*	*	*	*	*
GPM	(MLE)	1994 61	191 6	*	487 44	1975 61	*
J-M	(MLE)	*	2209 143	5981 418	3003 295	*	*
S-W	(MLE)	*	*	*	*	*	*
Binomial I	(LSE)	5175 180	*	*	*	***	***
Binomial II	(LSE)	*	*	*	*	***	***
IBM	(LSE)	14336 522	5133 340	13022 914	6742 663	***	***
IBM	(MLE)	4140 141	2331 151	5551 387	2681 263	***	***
No. Time Intervals		360	110	101	52	360	110

TABLE 5.3.14. CHI-SQUARE VALUES FOR DATA SET 14

Data Set Model		14-fd	14-fdG	14-fx	14-fxG	14-ff	14-ffG
GPM	(LSE)	716 41	*	483 82	118 27	711 41	*
J-M	(LSE)	1699 108	891 104	5595 957	585 136	1843 117	951 111
S-W	(LSE)	*	*	*	3641 856	*	*
GPM	(MLE)	701 40	*	426 72	119 28	709 41	*
J-M	(MLE)	945 57	476 54	4439 758	451 104	267 11	478 54
S-W	(MLE)	*	*	51492 8828	1818 426	*	52 2
Binomial I	(LSE)	1543 97	770 89	5011 857	507 117	***	***
Binomial II	(LSE)	1610 102	824 96	5226 893	543 126	***	***
IBM	(LSE)	1949 124	947 109	6547 1088	625 138	***	***
IBM	(MLE)	1020 62	493 56	4346 742	480 111	***	***
No. Time Intervals		111	36	19	11	111	36

TABLE 5.3.15. CHI-SQUARE VALUES FOR DATA SET 15

Data Set Model		15-fd	15-fdG	15A-fd	15A-fdG	15B-fd	15B-fdG	15C-fd	15C-fdG
GPM	(LSE)	538 35	58 4	269 21	*	*	*	298 24	28 2
J-M	(LSE)	1413 101	966 123	*	*	*	*	5268 507	2185 383
S-W	(LSE)	18785 1426	12950 1697	*	*	6104 1112	*	*	61579 10883
GPM	(MLE)	535 35	57 4	266 21	*	*	*	183 13	27 2
J-M	(MLE)	938 65	627 79	*	*	*	2 1	1452 136	1148 200
S-W	(MLE)	5485 412	4771 623	*	*	1585 287	*	16415 1589	22672 4005
Binomial I	(LSE)	*	*	*	*	*	*	5033 484	1967 345
Binomial II	(LSE)	*	*	*	*	*	*	*	2073 364
IBM	(LSE)	1389 99	958 120	4358043 419348	*	157 25	2 1	5169 492	2211 376
IBM	(MLE)	922 64	621 78	1276 119	*	*	2 1	1558 146	1199 209
No. Time Intervals		88	31	55	22	17	3	55	18

TABLE 5.3.16. CHI-SQUARE VALUES FOR DATA SET 16

Data Set Model		16-fd	16-fdG	16-fx	16-fxG	16-ff	16-ffG
GPM	(LSE)	3111 279	1920 268	869 14	56 -4	*	*
J-M	(LSE)	*	24123 3342	4334 128	1789 113	*	*
S-W	(LSE)	*	*	90676 2987	*	891591 99678	*
GPM	(MLE)	3147 352	1929 269	869 14	56 -4	*	*
J-M	(MLE)	10097 1124	10061 1392	1898 48	857 50	*	*
S-W	(MLE)	*	*	*	*	31 -1	0.03 -4
Binomial I	(LSE)	*	24048 3331	4265 12	1638 103	***	***
Binomial	(LSE)	*	*	4299 127	1533 107	***	***
IBM	(LSE)	19423 2140	23809 3236	4253 126	1854 116	***	***
IBM	(MLE)	9841 1096	9815 1357	1878 47	835 49	***	***
No. Time Intervals		42	28	458	113	42	28

TABLE 5.3.17. BEST FITS BY GPM, J-M, AND S-W MODELS

Model	Data Set	χ^2	Degrees of Freedom	Highest Level of Significance for Rejection of H_o*
GPM (LSE)	5-fdG	47	31	0.030
GPM (LSE)	5-fxG	26	33	0.80
GPM (MLE)	5-fdG	47	31	0.030
GPM (MLE)	5-fxG	26	33	0.80
S-W (MLE)	14-ffG	52	32	0.014
GPM (LSE)	15-fdG	58	28	0.00073
GPM (LSE)	15C-fdG	28	15	0.021
GPM (MLE)	15-fdG	57	28	0.0010
J-M (MLE)	15B-fdG	2	1	0.16
GPM (MLE)	15C-fdG	27	15	0.029
GPM (LSE)	16-fxG	56	110	>0.99999
GPM (MLE)	16-fxG	56	110	>0.99999
S-W (MLE)	16-ff	31	40	0.85
S-W (MLE)	16-ffG	.03	26	**
GPM (LSE)	5-ffG	47	31	0.030
GPM (MLE)	5-ffG	47	31	0.030

* I.e., H_O would not be rejected at any significance level below the values tabled here.

** Accurate value not available but practically equal to 1.

Indeed, the only other cases for which H_0 reasonably would not be rejected occurred for the IBM model in Table 5.3.15 and for the Binomial I in Table 5.3.12, and in all these cases the number of time intervals was only 3 leaving only one degree of freedom for the χ^2 statistic. Therefore, although no model is seen to fit consistently, the GPM is the most widely applicable.

Whether or not H_0 is true the χ^2 statistic is a measure of how closely the observed data is approximated by the expectations corresponding to the given model. In this sense, models and data sets can be compared. For example, the GPM, J-M, or S-W models quite often showed lower values of χ^2 over a fixed data set than the other models. In fact only in Tables 5.3.4 (command and control data), 5.3.6 (generalized information processing system), 5.3.9 (Ref (II,2)), and 5.3.12 (Ref (II,2)) were the GPM, J-M, or S-W outperformed. In all other cases (i.e., cases where fits were achieved with either the GPM, J-M, or S-W and the IBM and/or Binomial models) one of the models (GPM, J-M, or S-W) had the lowest χ^2 value over all other models for a fixed data set. Except for Table 5.3.12, the IBM model had the lowest χ^2 for cases where the GPM, J-M, and S-W were outperformed. In Table 5.3.12 (data set 12-fdG) the Binomial I was the best.

Finally, it should be pointed out that whenever both the MLE and LSE techniques yielded estimates for a given model, the χ^2 value associated with the MLE case is always smaller than the χ^2 value associated with the LSE case. This is not surprising, however, in view of the results discussed in Section 5.1.2. Furthermore, as discussed in Section 5.2, grouping the data such that there were 10 or more observations per debugging-time interval nearly always improved the χ^2 value (note that all but one of the best fits in Table 5.3.17 were for these regrouped data sets). Indeed, many of the extremely large normalized χ^2 values may be explained by observing that for all models (except possibly the GPM) the expected number of errors in a given debugging-time interval approaches zero as the time interval length approaches zero. Hence recording (perhaps erroneously) some errors in a very short time interval will have profound effects on the χ^2 statistic.

A possible, perhaps probable, reason for the poor performance of these S/W reliability models as demonstrated in Tables 5.3.1 through 5.3.16 is the violation of the model internal assumptions. For a further discussion of these considerations, see Section 6.0.

5.4 OPTIMUM TIME INTERVAL LENGTHS

In the models considered in this study the debugging-time interval lengths are not, strictly speaking, parameters but are part of the observations. Therefore, there is no freedom to "choose" them after the data have been collected. Guidelines may be set in choosing the debugging-time interval lengths before data is collected but this was not done for any of the data in this report.

Optimality of debugging-time interval lengths must be defined in terms of goodness of fit for a particular model. In this study, non-overlapping debugging-time intervals were combined so that the number of observations in each total time

interval was greater than or equal to 10 to ensure the validity of the distribution of the chi-square statistic used for testing goodness of fit (see Section 5.2). The results presented in Section 5.3 indicate that in most cases, this procedure improved the fit of the model. It should not, however, be concluded that further lengthening the debugging-time intervals will further improve the fit. Indeed, one could combine all the intervals into one large interval and obtain no fit since there would be, then, only one observation and two or more parameters to estimate. This does suggest, though, that an optimal time interval length exists but the problem of finding this optimal length is not well posed because a) this length will undoubtedly depend on the unknown parameters N, ϕ, α, a, b, etc. and b) estimates for the unknown parameters cannot be substituted because they will depend on the debugging-time interval lengths. Simulation techniques like those employed in Section 4.4 may be useful for fixed values of the unknown parameters, but would be extremely expensive and were not investigated in this study.

5.5 COMPARISON OF MODELS

5.5.1 Basic Model Similarities

Many basic similarities were seen among the models studied. Both the LS and ML methods of obtaining parameter estimates were appropriate for use with all the models. For the binomial models of Section 3.3, however, the ML method proved to be much too costly for practical use. Applying the Newton-Raphson method for each integer value of N in the input range, together with the evident need for a very large range, caused the computer runs to be both time consuming and expensive. This tends to imply the probable impracticality in testing only integer values of N in other models, as well.

Five of the models being considered give estimates for the parameter N, the total number of errors originally present. These models do not allow for the possible introduction of new errors (as corrections are made) during the software debugging process. The IBM model of Section 3.2, however, gives an estimate for $a = \lim_{t \to \infty} E\,(N(t))$, the total number of errors expected ever to occur. This is the only model which accommodates the possibility of introducing new errors while correcting others. However, because of the presence of this added factor, the unknown number of initial errors is not explicitly present in this model.

Due to this basic difference in the parameters N and a, inferences about the introduction of new errors during the debugging process can be made. If the estimates obtained for a and N are generally close, the assumption that no (or few) new errors are introduced would be strengthened. If, however, estimates for a are much larger than estimates for N, the presence of errors introduced in the debugging stage should be assumed. In actual testing, both cases were observed (see Table 5.5.1), although close estimates were more prevalent. This may indicate that the number of errors introduced through debugging is small relative to the number of errors present initially in the software.

TABLE 5.5.1. SELECTED MODEL COMPARISONS FOR THE ESTIMATION OF N

Data sets showing closeness and wide difference between estimates obtained for the model parameters N and a.

	Model	Data Set 7-fd	Data Set 16-fd
Estimates	GPM (LSE)	4720	2572
For	J-M (LSE)	4451	*
N	S-W (LSE)	3787	*
	GPM (MLE)	4618	1703
	J-M (MLE)	4458	3698
	S-W (MLE)	4186	*
	BIN I (LSE)	4450	*
	BIN II (LSE)	4451	*
Estimates	IBM	4230	1.22×10^8
For	IBM	4470	13826
a			

The GPM described in Section 3.1 requires estimates of three different parameters, while all other models contain only two. Estimating an increased number of parameters provides a larger freedom for the separation of different influences on the data sets. In this three-parameter case, three different factors influencing the occurrence of errors can be distinguished. However, numerical (tractability) complications are also added, and convergence of the estimates becomes even more difficult to obtain.

All the models assume decreasing error rates. Although this assumption is made, very few of the actual data sets have uniformly decreasing error rates. This may have caused convergence problems in the application of the Newton-Raphson method.

5.5.2 Similarities Obtained for Different Models and Methods

Many inherent similarities among the models have been found. Most generally, the ML and LS methods for obtaining estimates usually resulted in similar estimates for the same model. This trend supports the validity of both methods. Most noticeably, extremely close estimates of ϕ were obtained through both methods for each of the Poisson type models of Section 3.1. The costliness of the ML methods for the binomial models of Section 3.3 is the only reason a comparison cannot be made in those cases.

For all the models, the grouped data sets had slightly better convergence properties (convergence was obtained more often and more easily) than the corresponding non-grouped data sets. In the majority of cases, grouping altered the parameter estimates obtained very little, although generally better chi-square values resulted for these fits. A direct comparison of these chi-square values for one data set is given in Table 5.5.2.

Extremely close values of the estimates of $\tilde{a}$ (in the LS version of the binomial model of Section 3.3) with parameter $p_i = 1 - e^{-a\tau_i}$ and $\tilde{\phi}$ (in the LS version of the Jelinski-Moranda model described in Section 3.1) were observed when both had been found. This can be explained by observing the least squares sums to be minimized.

In the J-M model it is

$$S_1 = \sum_{i=1}^{k} \left\{ N_i - (N - M_{i-1}) \phi \tau_i \right\}^2 ,$$

while in the binomial model it is

$$S_2 = \sum_{i=1}^{k} \left\{ N_i - (N - M_{i-1}) (1 - e^{-a\tau_i}) \right\}^2$$

Notice that the only difference occurs between the values $\phi\tau_i$ and $1 - e^{-a\tau_i}$. Series expansion shows that

$$1 - e^{-a\tau_i} = 1 - (1 - a\tau_i + \frac{(a\tau_i)^2}{2!} - \frac{(a\tau_i)^3}{3!} + \ldots)$$

$$= a\tau_i - \frac{(a\tau_i)^2}{2!} + \frac{(a\tau_i)^3}{3!} - \ldots .$$

TABLE 5.5.2. SELECTED PARAMETER ESTIMATES AND CHI-SQUARE RESULTS FOR ALL MODELS

		Estimates		Chi-Square Values Normalized Chi-Square Values	
Model		14-fx	14-fxG	14-fx	14-fxG
	$\tilde{N}$	541	3972	483	118
GPM	$\tilde{\alpha}$	-.3404	-.2314	82	27
	$\tilde{\phi}$	.1876	.0223		
J-M	$\tilde{N}$	609	793	5595	585
	$\tilde{\phi}$	.0055	.0043	957	136
S-W		*	633	*	3641
			2.07×10^{-4}		856
	$\hat{N}$	555	2727	426	119
GPM	$\hat{\alpha}$	-.1605	-.1200	72	28
	ϕ	.1392	.0251		
J-M	$\hat{N}$	773	763	4439	451
	$\hat{\phi}$	.0055	.0056	758	104
S-W	$\hat{N}$	693	593	51492	1818
	ϕ	.0043	4.36×10^{-4}	8828	426
Bin. I	$\tilde{N}$	618	763	5011	507
	$\tilde{c}$	167	193.78	857	117
Bin. II	$\tilde{N}$	614	772	5226	543
	$\tilde{a}$	.0058	.0047	893	126
IBM	$\tilde{a}$	471	656	6547	625
	$\tilde{b}$	2138	.1577	1088	138
IBM	$\hat{a}$	996	842	4346	480
	$\hat{b}$	.1169	.1511	742	111

The portion of the series, $-\frac{(a\tau_i)^2}{2!} + \frac{(a\tau_i)^3}{3!} + - \ldots$, is approximately zero for small values of $a\tau_i$.

This gives

$$S_2 \cong \sum_{i=1}^{k} \left\{ N_i - (N - M_{i-1})\, a\tau_i \right\}^2$$

which is exactly S_1 if ϕ is substituted for a (similar analysis will lead to the same result for the Binomial model with $p_i = \tau_i/(\tau_i + c)$ for $\phi = 1/c$ and c large).

Therefore, since the values found for both $\tilde{\phi}$ and $\tilde{a}$ are usually small, their closeness is to be expected.

5.6 RECOMMENDED MODELS

Basically, the results of the goodness-of-fit tests and other aspects of the data analysis indicate none of the models fit very well. However, as discussed in Section 6.0, the various data sets showed obvious tendencies to increasing (in time) error rates: more or less contrary to the model assumption that mean error rates should be non-increasing. Also it seems clear that the observed increasing error rate situation cannot simply be written off as due the introduction of errors during the debugging process or as due purely to chance: the fluctuations are too large. Thus the data sets themselves have some problems of a nature unknown to us; probably lack of constancy of the debugging effort both in terms of man power and the S/W modules available for test. Thus with better data more fits might have been obtained. The "proof" of this allegation will be given shortly.

It is not surprising that the Poisson models were the best fitting for they incorporate not only finds (N_i) but fixes (M_{i-1}). Thus finds do not have to be fixed immediately. It is also not surprising that the GPM model was the best fitting of Poisson models: it has 3-parameters and hence a great deal of flexibility. A possibly disturbing feature of the GPM results is that the estimates of shaping parameter α virtually always came out negative but near, very near, zero. When the estimates of α came out positive they were also near zero. This indicates that $\alpha = 0$ and hence the Poisson mean is of the approximate form $\phi\ (N - M_{i-1})$ since $\tau_i 0 = 1$. **This (the data) says that the error rate does not depend on τ_i much but it depends on other factors extant when the data were generated, like** non-constant debugging effort. Thus there is every indication that the failure to obtain a large number of fits was not the fault of the models above.

In view of the above we believe we can recommend the GPM and to a lesser extent the J-M and S-W models. The IBM model is unsatisfactory for, among other reasons, the reason that there is no parameter N in the model.

Section 6.0

VALIDATION OF INTERNAL ASSUMPTIONS

The crux of the matter, the thing that essentially determines how well a model fits the observed data, is are the assumptions explicit or implicit in the model valid for the stochastic process which actually generated the data? Another, related point, is how sensitive are the models to a failure (in the data) of the model assumptions to hold?

The S/W reliability models studied in this report have no lack of explicit and implicit assumptions.

Perhaps the foremost assumption in all of the models is that the expected (mean) value of the number of errors occurring in interval i, $i = 1, 2, \ldots, k$ (say) is a never increasing function of increasing i (provided of course account is taken of the possibly different magnitudes of the τ_i. For example in the S-W model (the GPM with $\alpha = 2$) $E(N_i) = \phi(N - M_{i-1})\tau_i^2$, and since M_{i-1} is never decreasing in i and assuming $\tau_1 = \tau_2 = \ldots = \tau_k$, clearly $E(N_i)$ is never increasing in i. The IBM model has a similar, perhaps more implicit assumption. By definition for the IBM model

$$E\,[\text{number of errors in } (0,\ t)] = a\,(1 - e^{-bt}).$$

Now consider an interval $(t,\ t+h)$, $t,\ h > 0$; i.e., an interval of width h.

$$E\,[\text{number of errors in } (t,\ t+h)]$$

$$= E\,[\text{number of errors in } (0,\ t+h)] - E\,[\text{number of errors in } (0,\ t)]$$

$$= a(1-e^{-b(t+h)}) - a(1-e^{-bt}) = ae^{-bt}\,(1-e^{-bh})$$

For a, b > 0 as required by the IBM model the function on the right in the last equation, for fixed h, is decreasing in t. A similar result is true for the Binomial models.

This assumption is extremely plausible because as errors are observed and corrected less errors remain. However, as plausible as this assumption is it has associated with it two serious questions. First, what happens if the assumption is false (perhaps due to increased number of debugging personnel in the later stages of S/W development)? Second, what happens if the assumption is valid but due to pure

random fluctuations the observed number of errors (as against the expected number) is increasing (at least part of the time) in i ? Regarding the first question it is simply a case of the model not being descriptive of the physical (stochastic) process generating the data. We feel that the failure of this assumption to be valid is a major (but not the only) cause of the failure of the models to fit the data with any reasonable frequency. The second question has actually more serious implications. This question is pointed toward the sensitivity of the models to the data. After all even if the expected number of errors is decreasing per unit time, purely random (e.g., according to the Poisson distribution for the GPM) fluctuations will cause some intervals to have more errors observed than previous intervals. If the models are over sensitive to this condition they are, in a very real sense, useless. Now quite apart from questions of goodness-of-fit of the models, we found (as discussed in Section 5.1) many cases of failure of the (numerical) parameter estimation procedure to converge or the (numerical) parameter estimation procedure led to absurd results (e.g., an estimate of N, the initial number of errors, smaller than the observed number of errors already removed). Even though the goodness-of-fit results (see Section 5.3) were mostly quite bad as measured by the observed standardized chi-square values it seems that the models are very sensitive to the failure of the data to show an observed decreasing (in time) error rate. This is indeed a severe shortcoming. To further emphasize this problem suppose that a chi-square goodness-of-fit test is run on some data on the hypothesis that the data follows a normal distribution. If the data really are from, say a gamma distribution, (i.e., the hypothesized model (normal) is incorrect) the goodness-of-fit procedure at least does not break down! That is it won't fail to give some kind of answer. Apparently, when the observed data depart from the deterministic (decreasing mean error rate) assumptions, for whatever reason, the models often blow-up in the sense that no parameter estimates are even obtainable. To check the behavior of the observed error rate we fit parabolas and straight lines to the observed error rate versus time for a number of data sets. Some of the data sets were omitted: the -G (grouped) data sets were omitted because they were not the original data and the ff data sets were not used because they have no meaning in this calculation.

The straight lines and parabolas, whose equations are given in Table 6.1, were not fitted for any predictive purposes; they were fitted solely to try to obtain some single or overall measure of the behavior of the error rate versus time. The results are depicted in Figures 6.1.1 - 6.1.27. About one-quarter (7/27) of the straight lines showed positive slope indicating, admittedly very roughly, an increasing (in time) error rate. Many of the parabolas reach the maximum (and hence start down) relatively far into the time frame. Moreover some of the parabolas had minimums instead of maximums and the error rate started back up late in time. On inspecting Tables 5.1.1 - 5.1.16 it can be seen that quite severe (except for the IBM model) convergence problems were experienced for data sets 1-fd, 3-fd, 3-fx and 13-fx. Inspecting the corresponding Figures 1, 4, 5 and 19 the straight lines had positive slope in 3 of the 4 cases (the fourth case had a slope not all that different from zero). All four cases had parabolas reaching their maximum late in time; hence the error rate started down very far into the program. This would seem to give a general indication that when the data "misbehaved," albeit due to purely random fluctuations, the models "blew-up." Of course whatever the source of the observed increasing error rate the models (except for the IBM model) were very sensitive to it. The IBM model had its own problems; for the same data sets (1-fd, 3-fd, 3-fx, 13-fx) the IBM model gave absurdly high estimates of a (for 3-fd and the MLE $\hat{a}$ was almost 10^{13} in the face of 366 observed errors). For 3-fx, convergence for a was not even obtained. For data set 13-fx $\tilde{a}$ = (937) was below the observed errors (1789).

TABLE 6.1. ERROR RATE VERSUS CUMULATIVE TIME EQUATIONS
(x ≡ cumulative elapsed time, dependent variable ≡ errors/day)

Data Set No.	Parabola	Straight Line
1-fd	$-4.7E{-5}x^2 + 4.1E{-2}x - 2.5$	$9.1E{-3}x + 1.8$
1-fx	$-6.3E{-5}x^2 + 5.9E{-2}x - 4.2$	$1.6E{-2}x + 0.91$
7-fd	$3.8E{-5}x^2 - 7.0E{-2}x + 21.6$	$-5.9E{-2}x + 21.0$
7-fx	$2.1E{-4}x^2 - 7.6E{-2}x + 14.7$	$1.9E{-2}x + 7.4$
2-fd	$1.9E{-5}x^2 - 3.7E{-2}x + 20.5$	$-1.7E{-2}x + 16.9$
3-fd	$-2.8E{-4}x^2 + 7.0E{-2}x - 0.27$	$1.2E{-2}x + 2.3$
3-fx	$-2.9E{-4}x^2 + 7.1E{-2}x + 2.5E{-2}$	$1.4E{-2}x + 2.3$
5-fd	$-6.5E{-4}x^2 + 1.4E{-1}x - 1.5$	$-6.6E{-3}x + 4.6$
5-fx	$-1.3E{-4}x^2 + 3.7E{-2}x + 1.2$	$-8.1E{-3}x + 4.3$
15C-fd	$2.2E{-7}x^2 - 8.5E{-4}x + 0.90$	$-3.5E{-4}x + 0.70$
15A-fd	$-2.7E{-7}x^2 + 6.5E{-4}x - 1.1E{-2}$	$-4.3E{-5}x + 0.35$
15B-fd	$6.4E{-9}x^2 - 3.5E{-5}x + 7.1E{-2}$	$-2.0E{-5}x + 6.6E{-2}$
6-fd	$-1.8E{-5}x^2 + 1.2E{-2}x + 0.63$	$4.5E{-3}x + 1.1$
8-fd	$1.2x^2 - 28.6x + 202.9$	$-8.9x + 147.1$
9-fd	$1.3x^2 - 32.1x + 244.4$	$-10.6x + 183.5$
10-fd	$2.4x^2 - 51.6x + 395.5$	$-13.4x + 287.5$
11-fd	$2.3x^2 - 48.6x + 356.6$	$-11.9x + 25.2$
12-fd	$-7.3E6x^2 + 3.0E{-3}x + 0.16$	$10.0E{-4}x + 0.25$
15-fd	$-1.4E{-7}x^2 + 3.4E{-4}x + 0.18$	$2.8E{-5}x + 0.30$
14-fd	$-6.2E{-6}x^2 - 1.9E{-2}x + 5.3$	$-2.0x + 5.4$
14-fx	$-6.7E{-4}x^2 + 7.0E{-2}x + 11.0$	$-6.4E{-2}x + 15.6$
13-fd	$3.2E{-6}x^2 - 8.4E{-3}x + 5.6$	$-5.1E{-3}x + 5.2$
13-fx	$-4.3E{-6}x^2 - 1.3E{-3}x + 6.3$	$-5.6E{-3}x + 6.8$
4-fd	$-2.5E{-3}x^2 + 3.5E{-1}x + 19.6$	$-7.7E{-2}x + 31.9$
4-fx	$-5.9E{-4}x^2 + 7.8E{-2}x + 25.8$	$-6.4E{-2}x + 31.8$
16-fd	$-3.6E{-5}x^2 + 2.4E{-2}x + 3.4$	$-5.7E{-3}x + 7.2$
16-fx	$-1.2E{-5}x^2 + 6.3E{-3}x + 2.1$	$-2.1E{-3}x + 3.2$

NOTE: $EY = 10^Y$

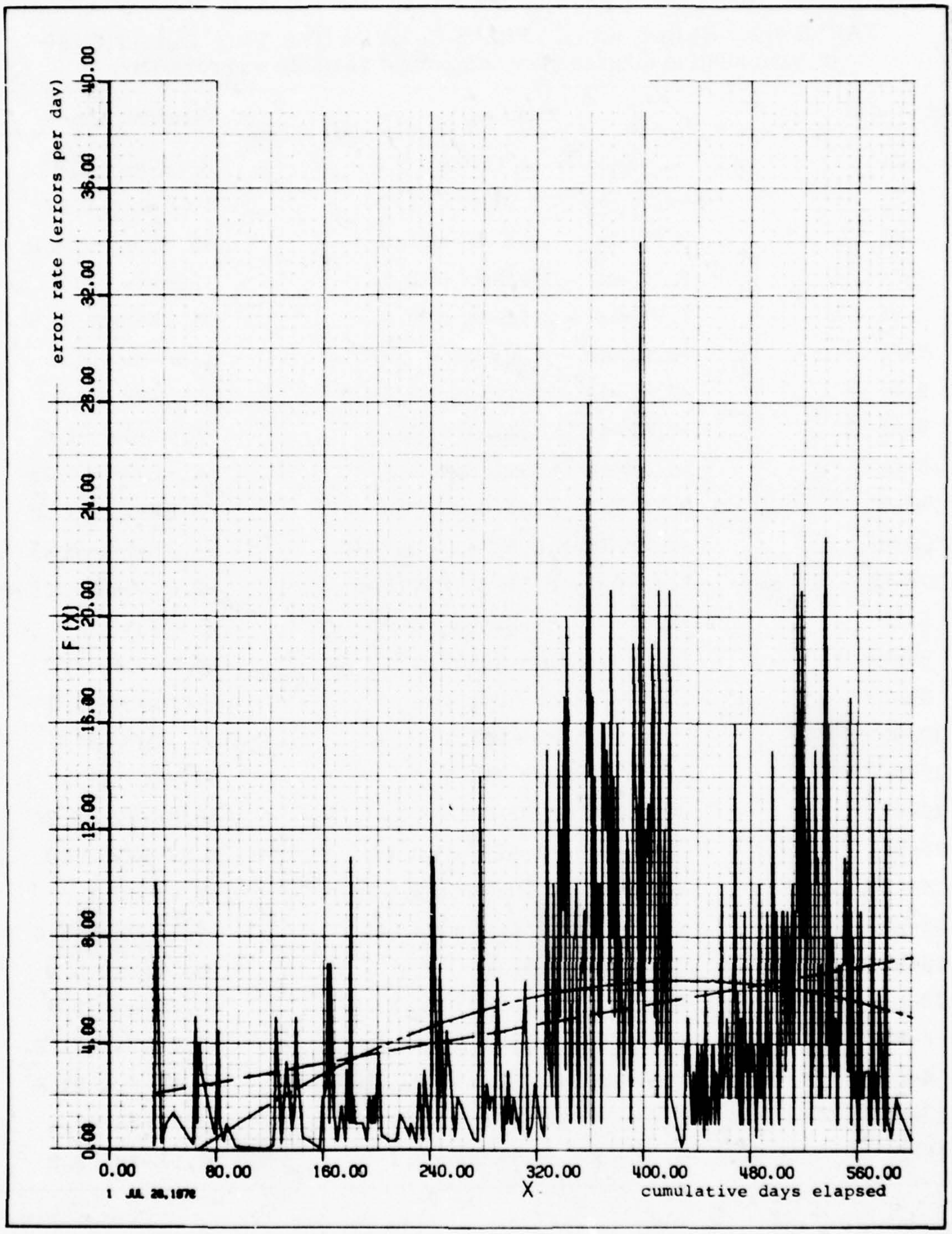

Figure 6.1.1. Error Rate Versus Cumulative Time – Data Set 1-fd

82755-44

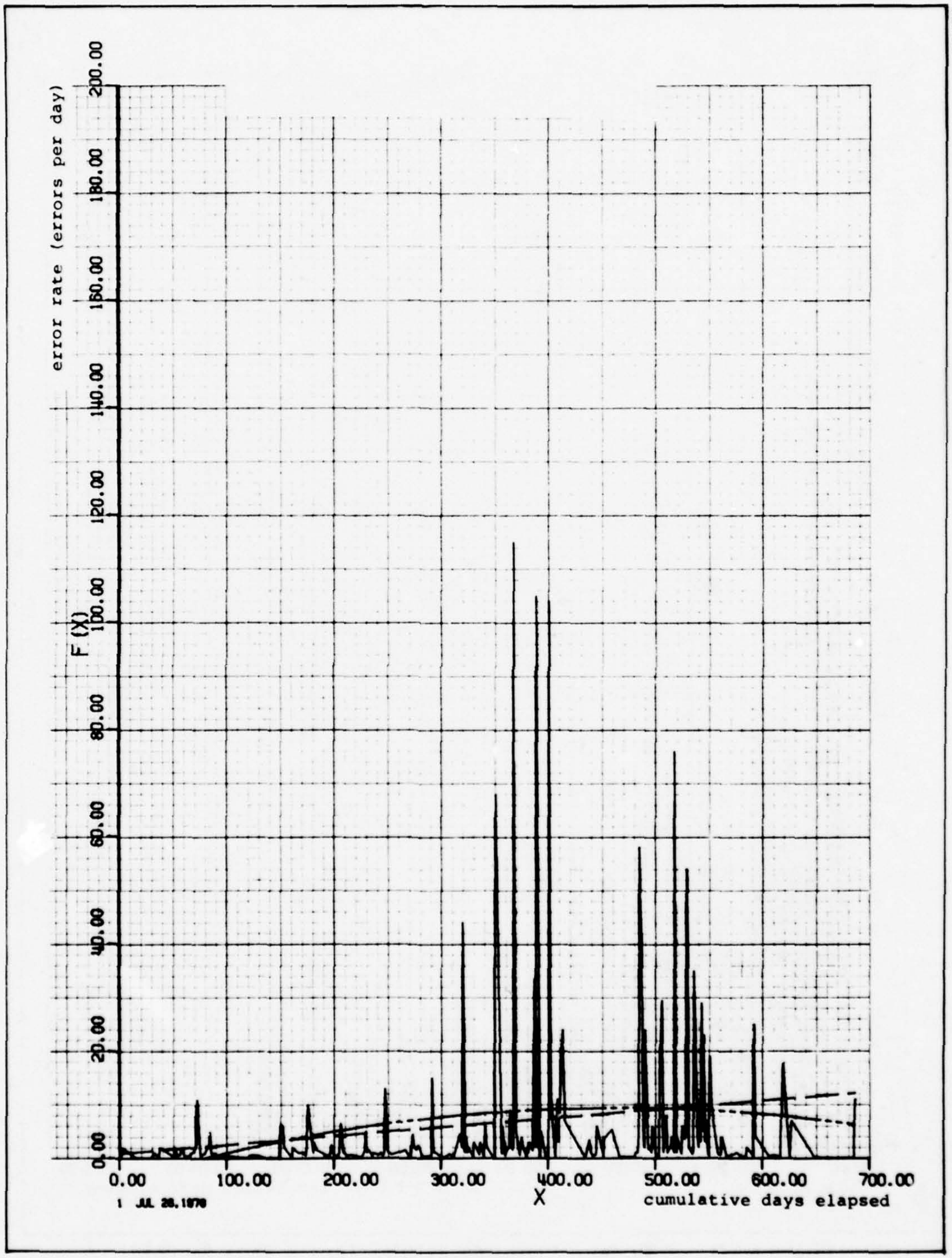

Figure 6.1.2. Error Rate Versus Cumulative Time – Data Set 1-fx

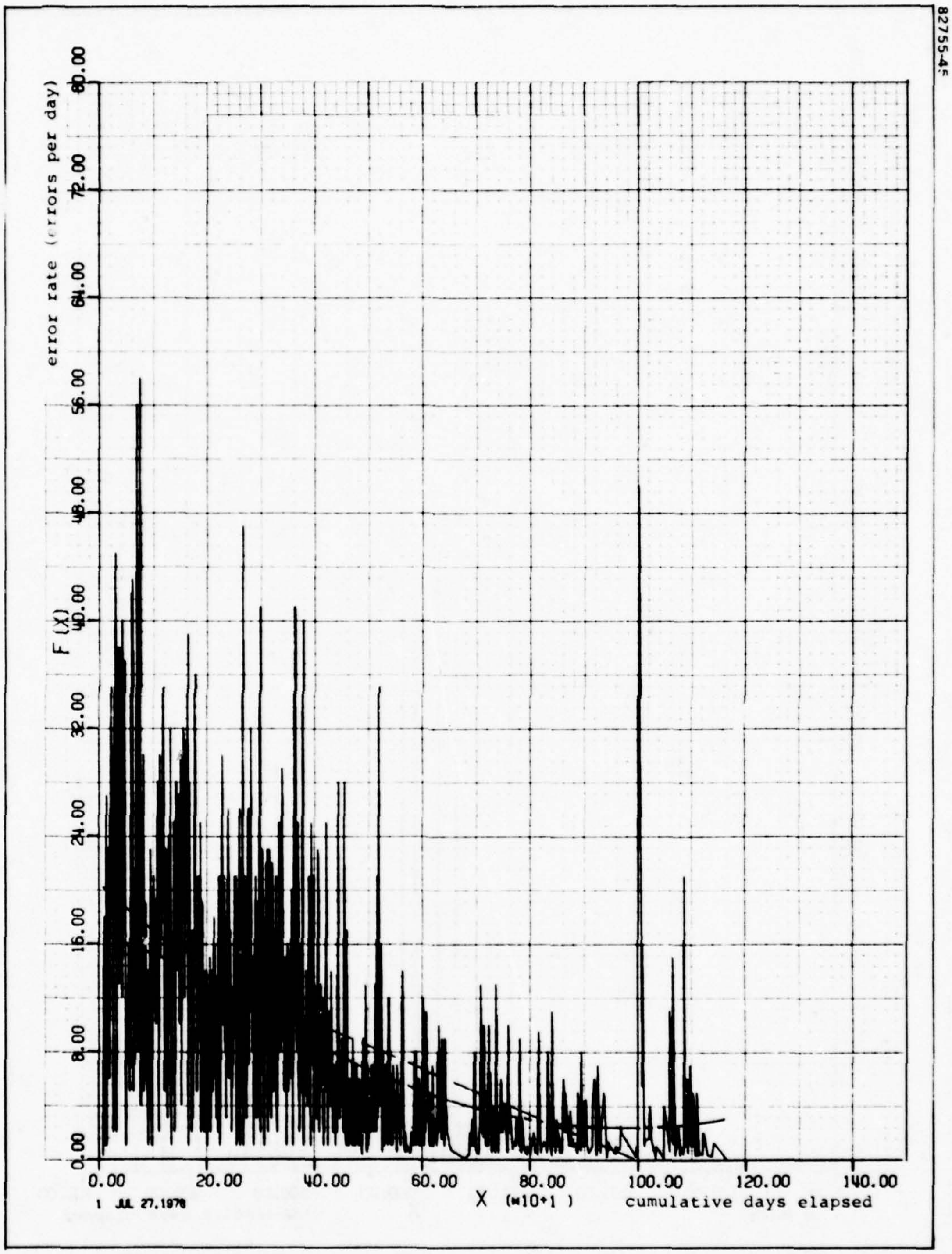

Figure 6.1.3. Error Rate Versus Cumulative Time – Data Set 2-fd

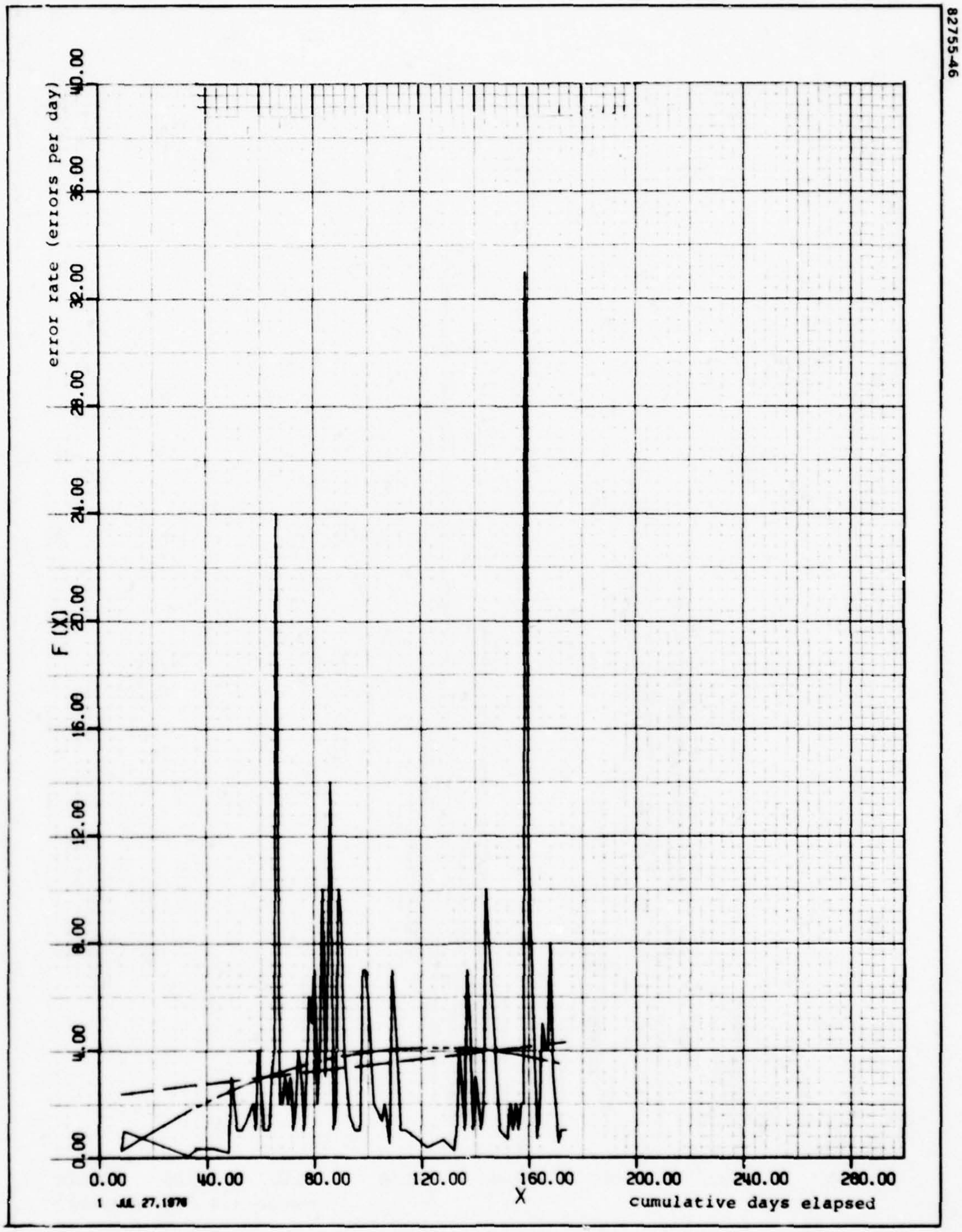

Figure 6.1.4. Error Rate Versus Cumulative Time – Data Set 3-fd

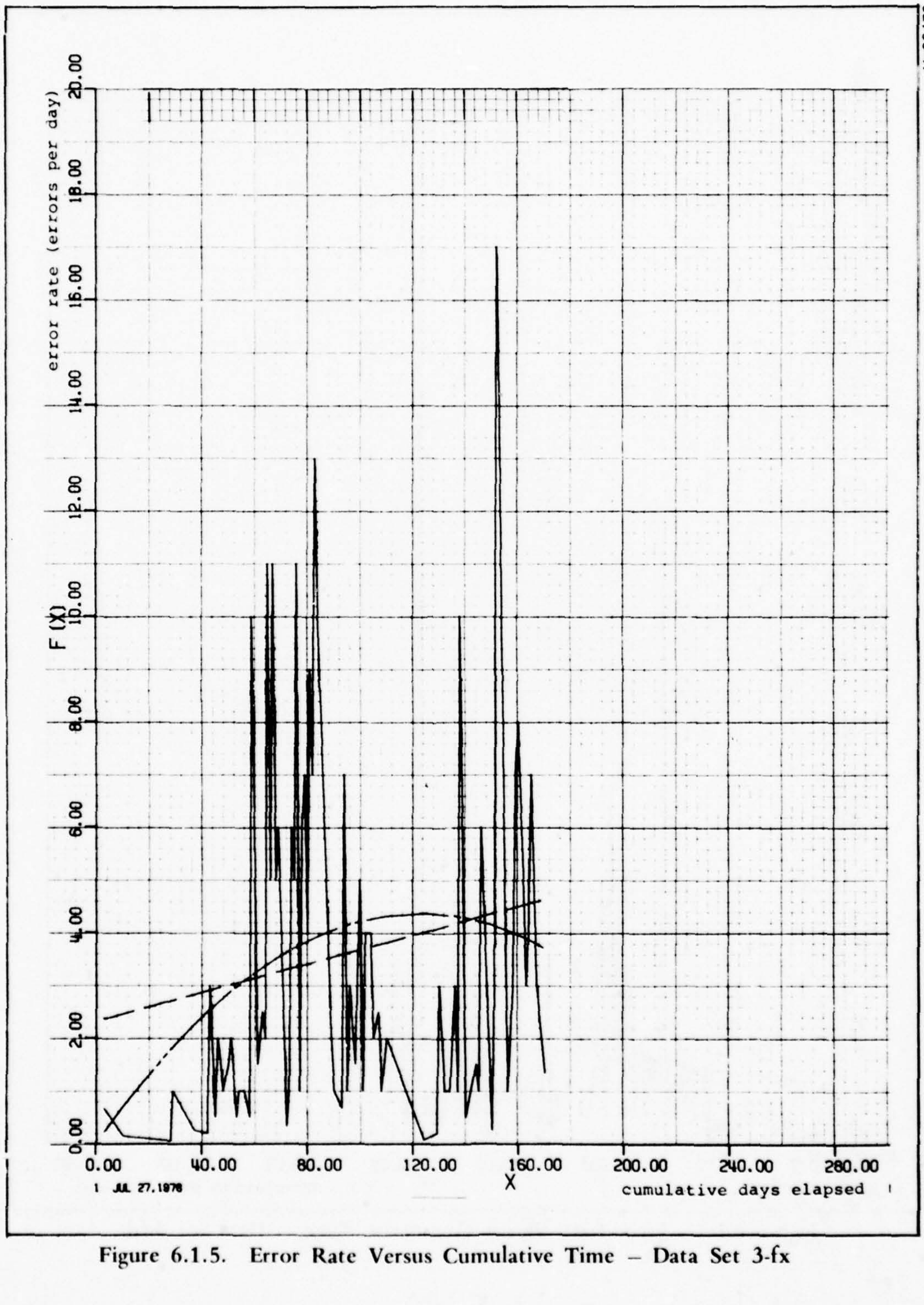

Figure 6.1.5. Error Rate Versus Cumulative Time – Data Set 3-fx

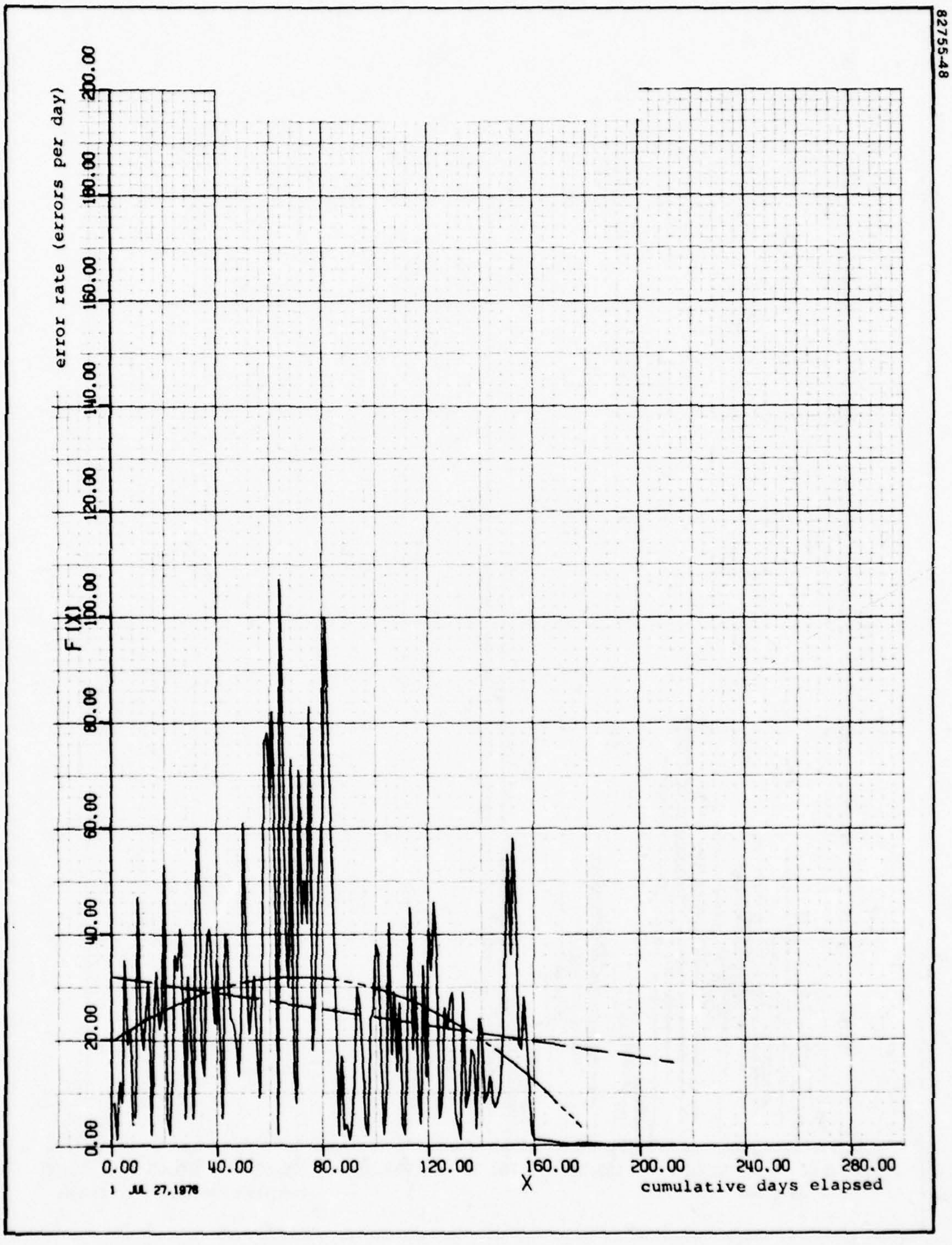

Figure 6.1.6. Error Rate Versus Cumulative Time – Data Set 4-fd

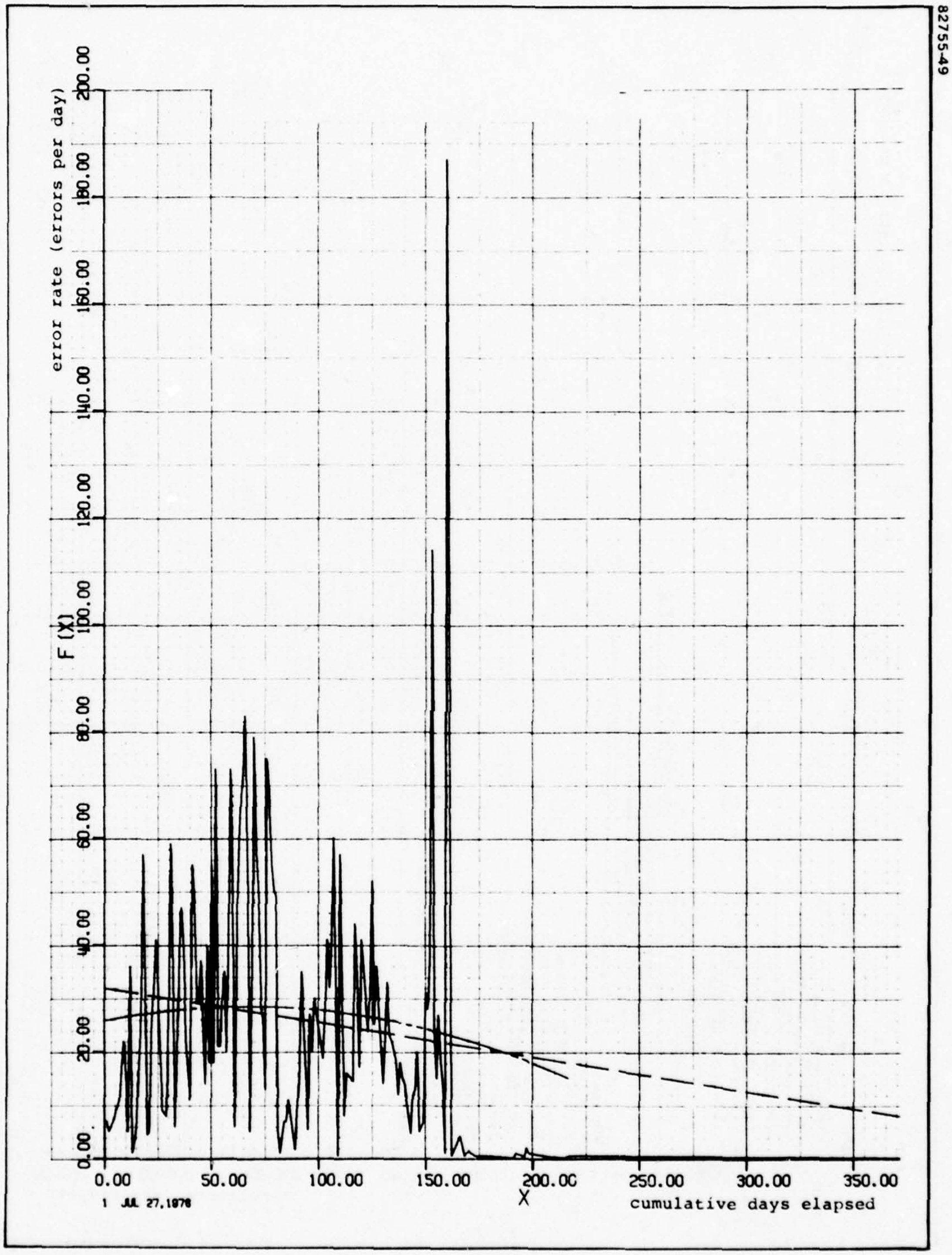

Figure 6.1.7. Error Rate Versus Cumulative Time – Data Set 4-fx

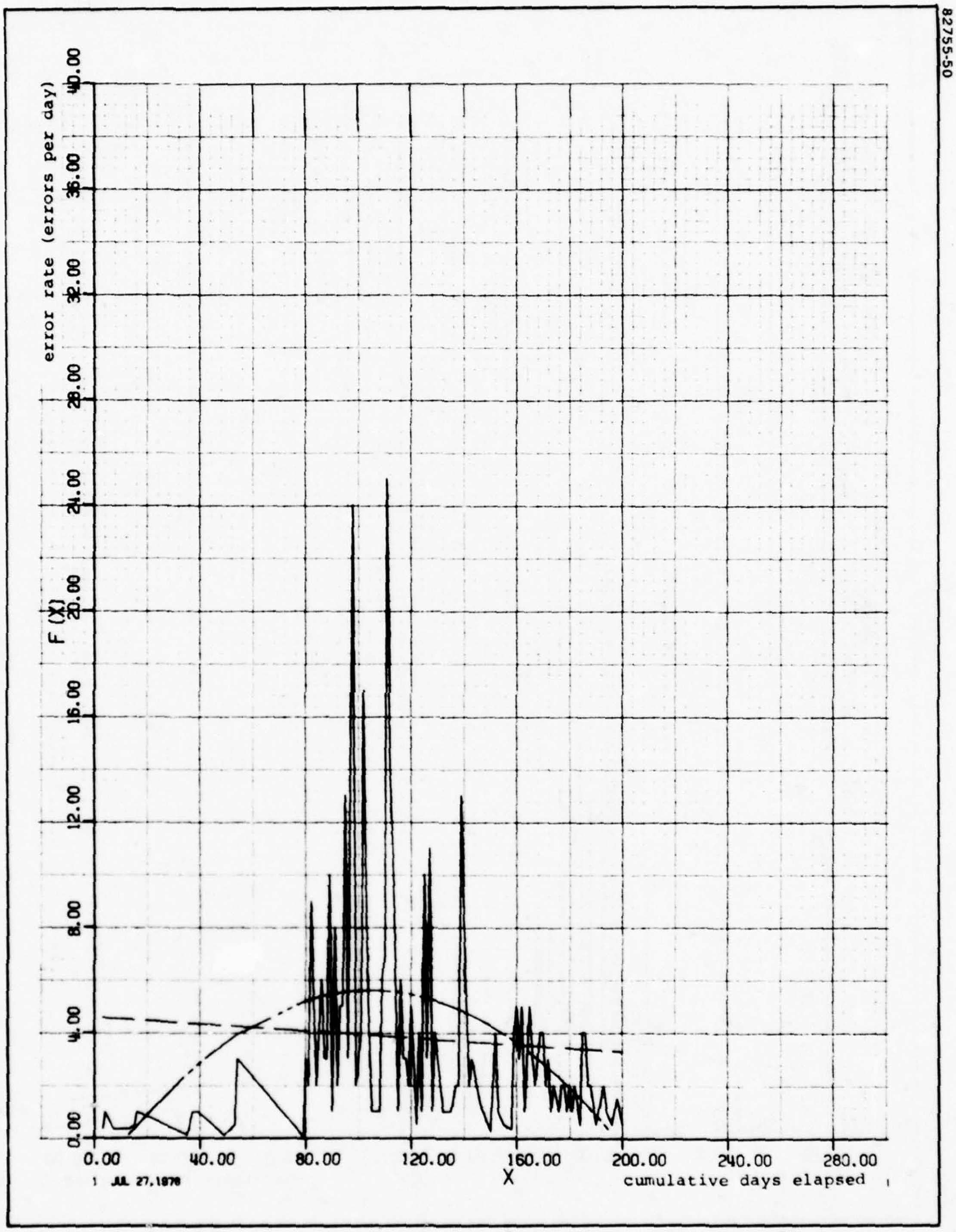

Figure 6.1.8. Error Rate Versus Cumulative Time – Data Set 5-fd

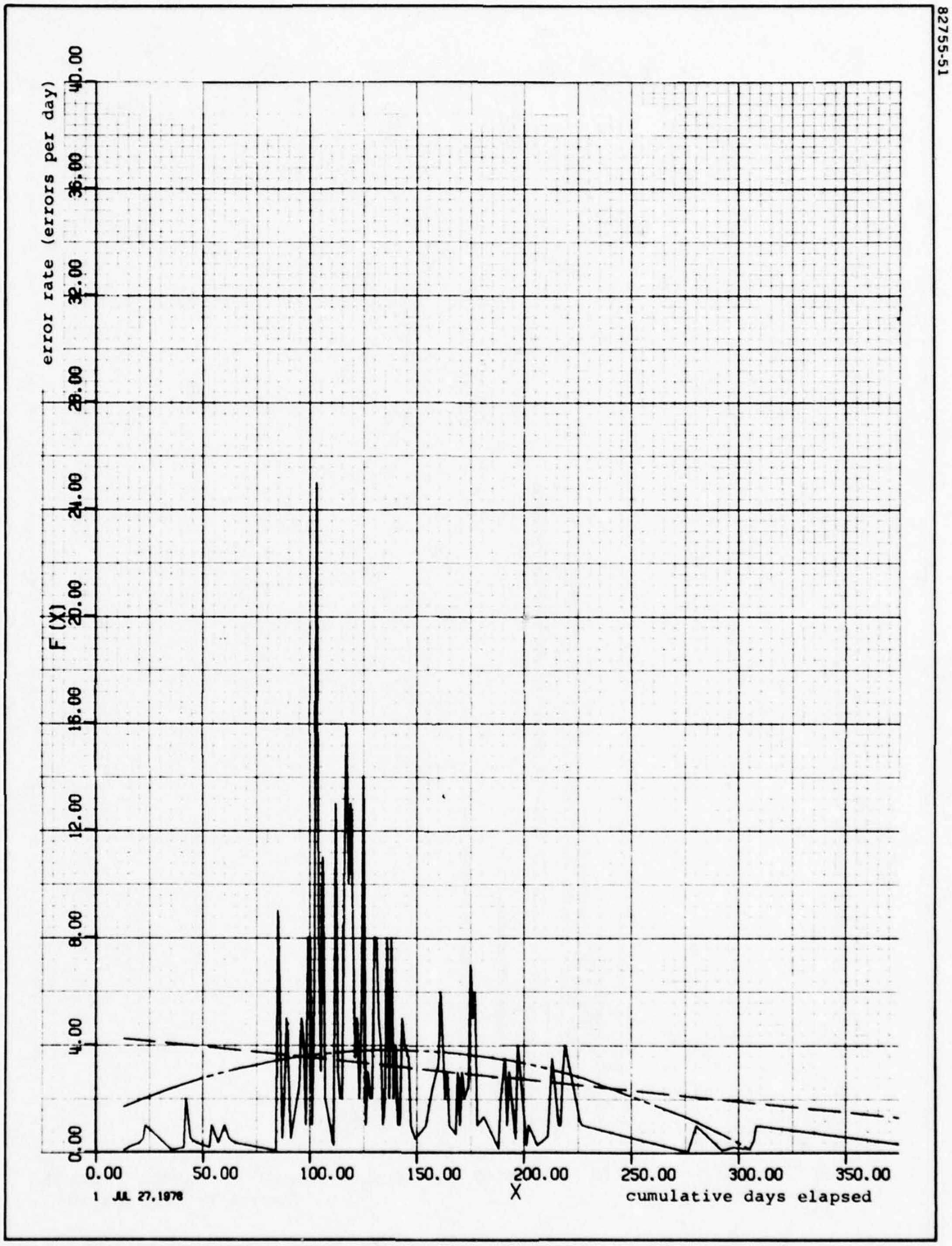

Figure 6.1.9. Error Rate Versus Cumulative Time – Data Set 5-fx

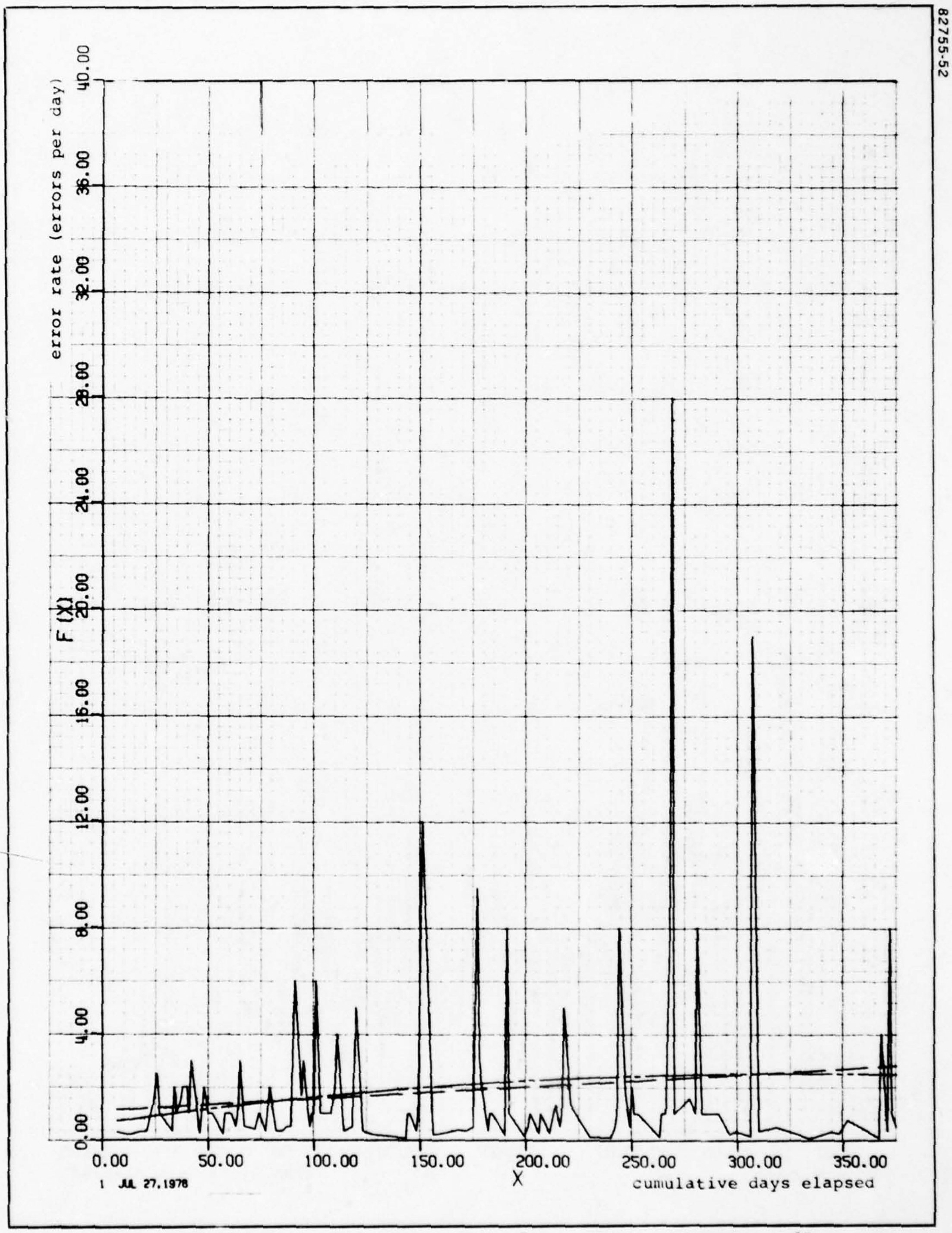

Figure 6.1.10. Error Rate Versus Cumulative Time – Data Set 6-fd

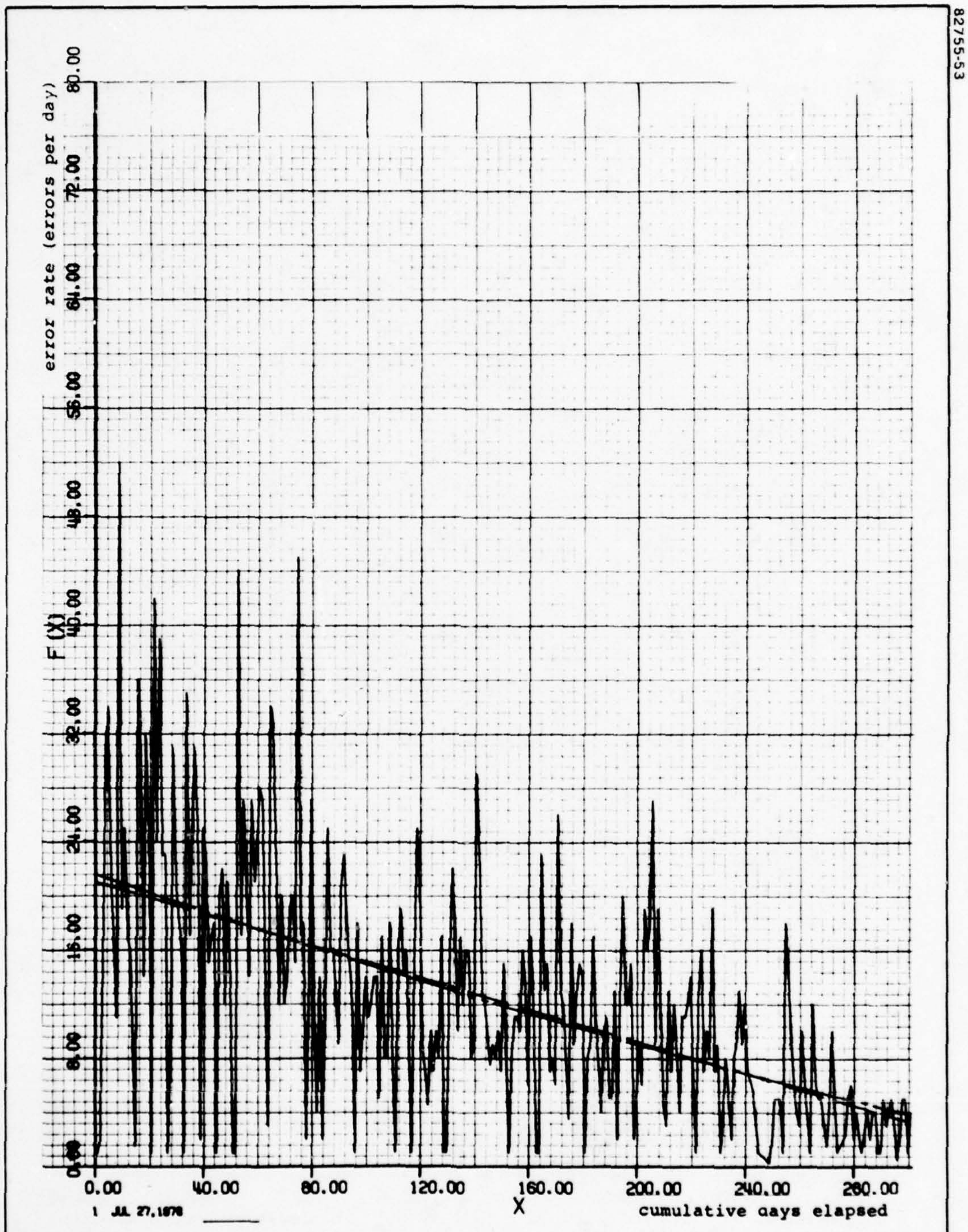

Figure 6.1.11. Error Rate Versus Cumulative Time – Data Set 7-fd

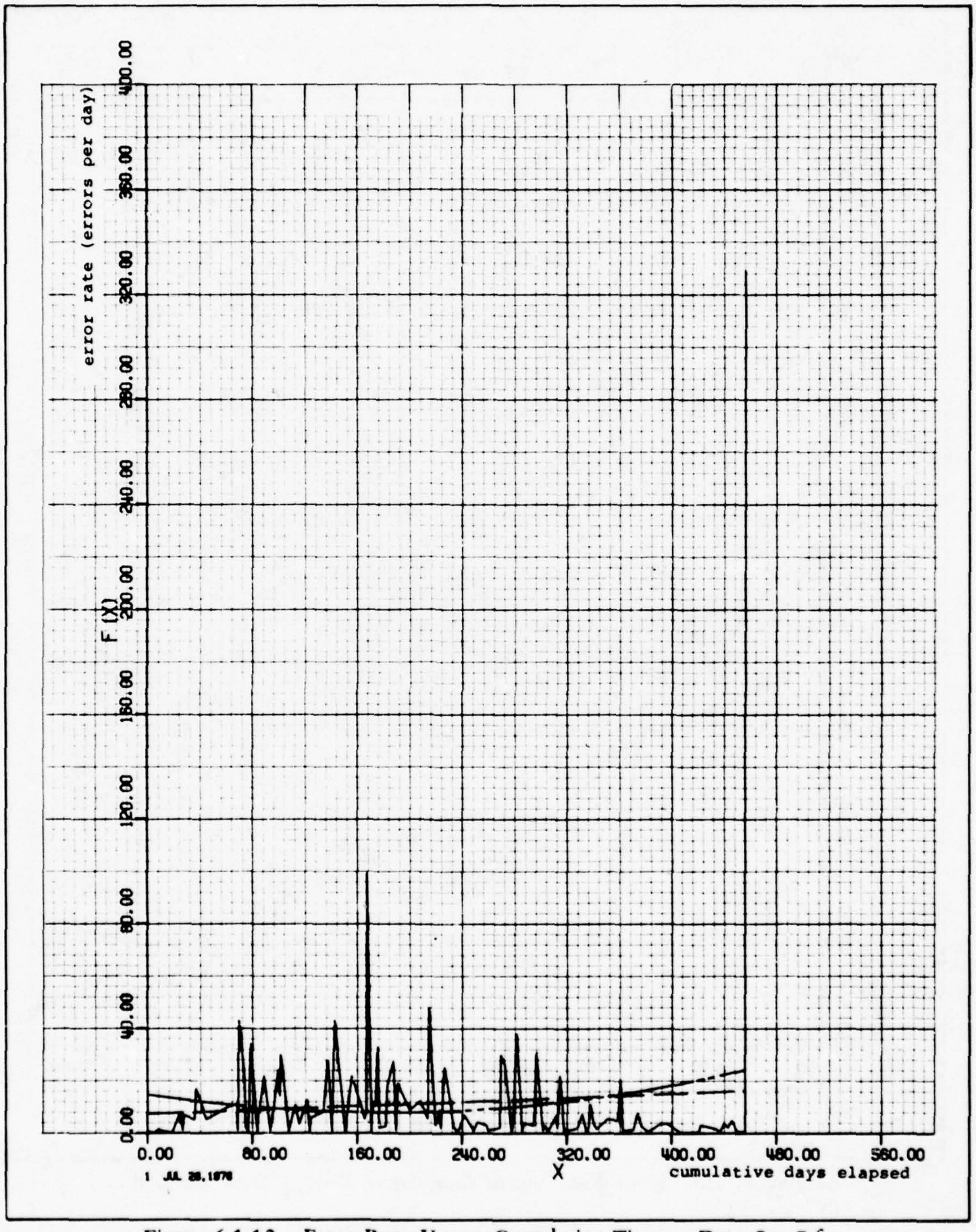

Figure 6.1.12. Error Rate Versus Cumulative Time – Data Set 7-fx

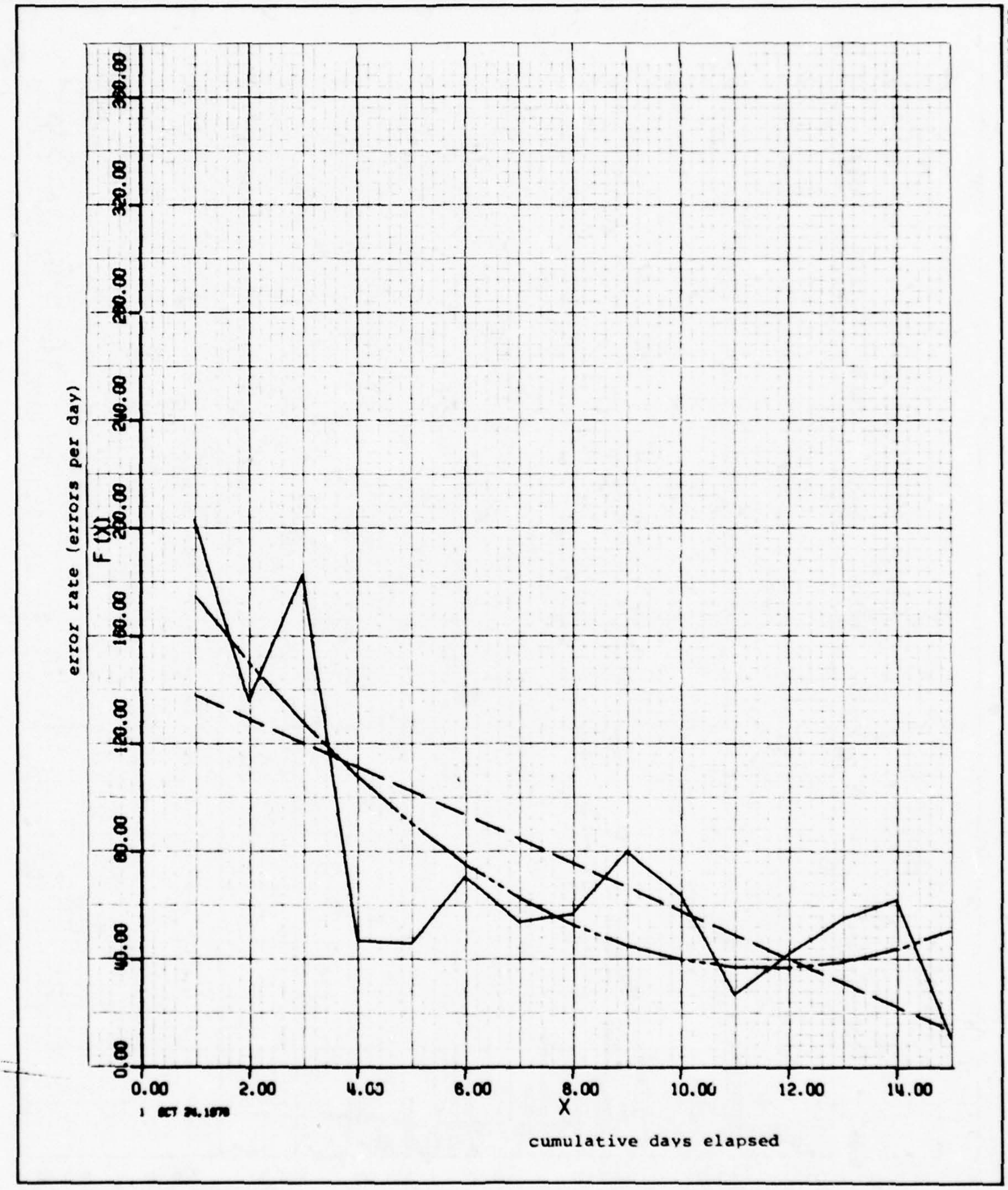

Figure 6.1.13. Error Rate Versus Cumulative Time – Data Set 8-fd

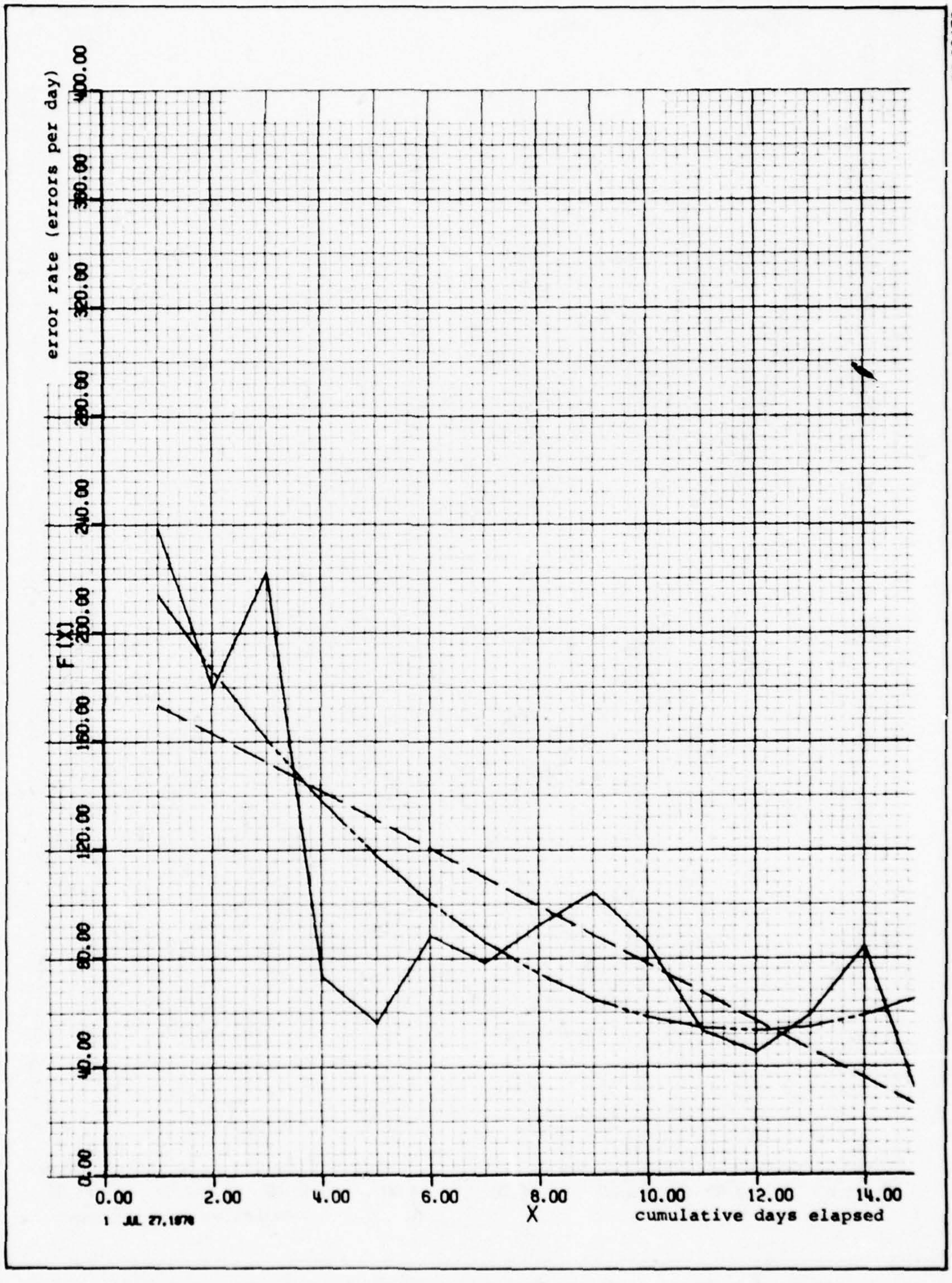

Figure 6.1.14. Error Rate Versus Cumulative Time – Data Set 9-fd

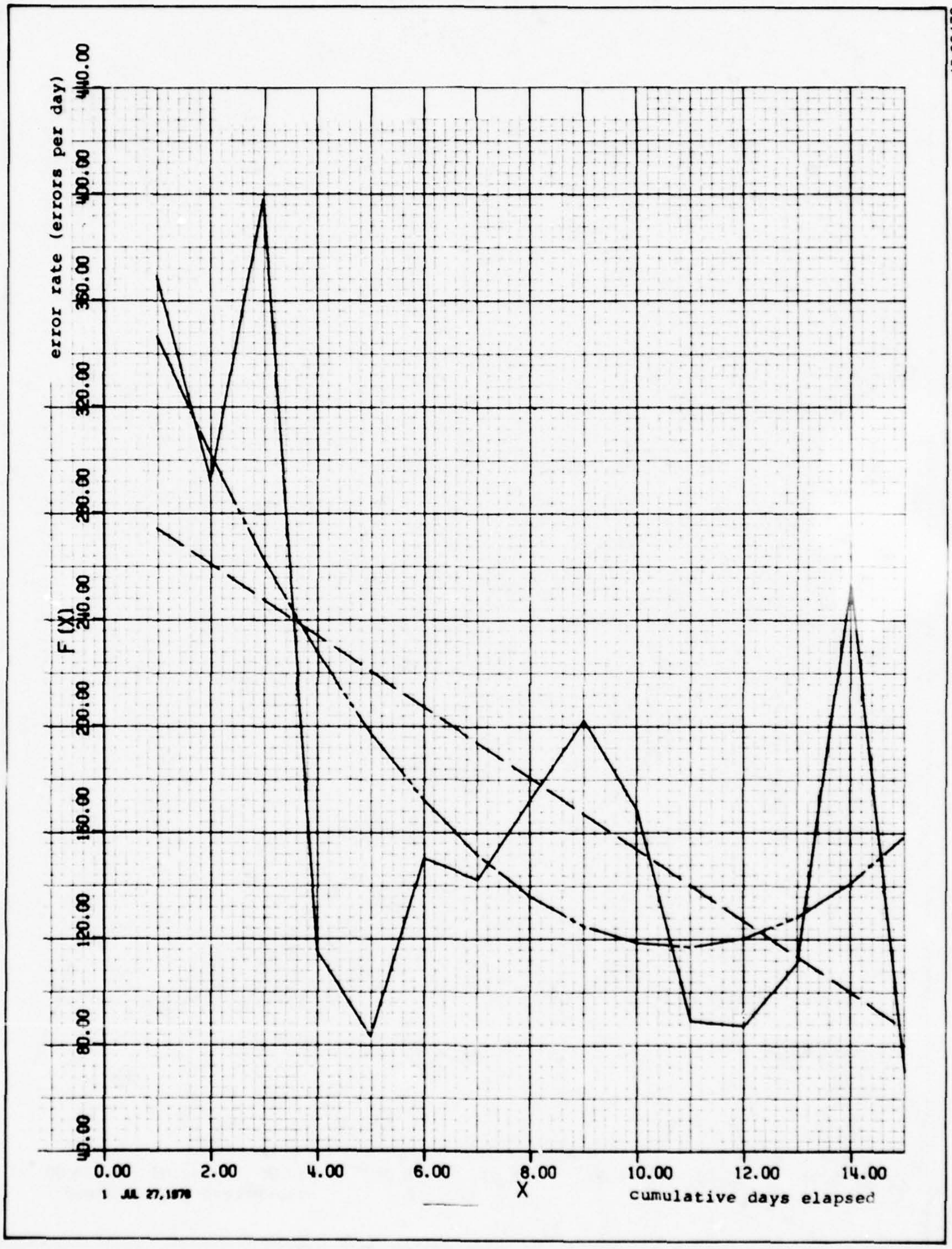

Figure 6.1.15. Error Rate Versus Cumulative Time – Data Set 10-fd

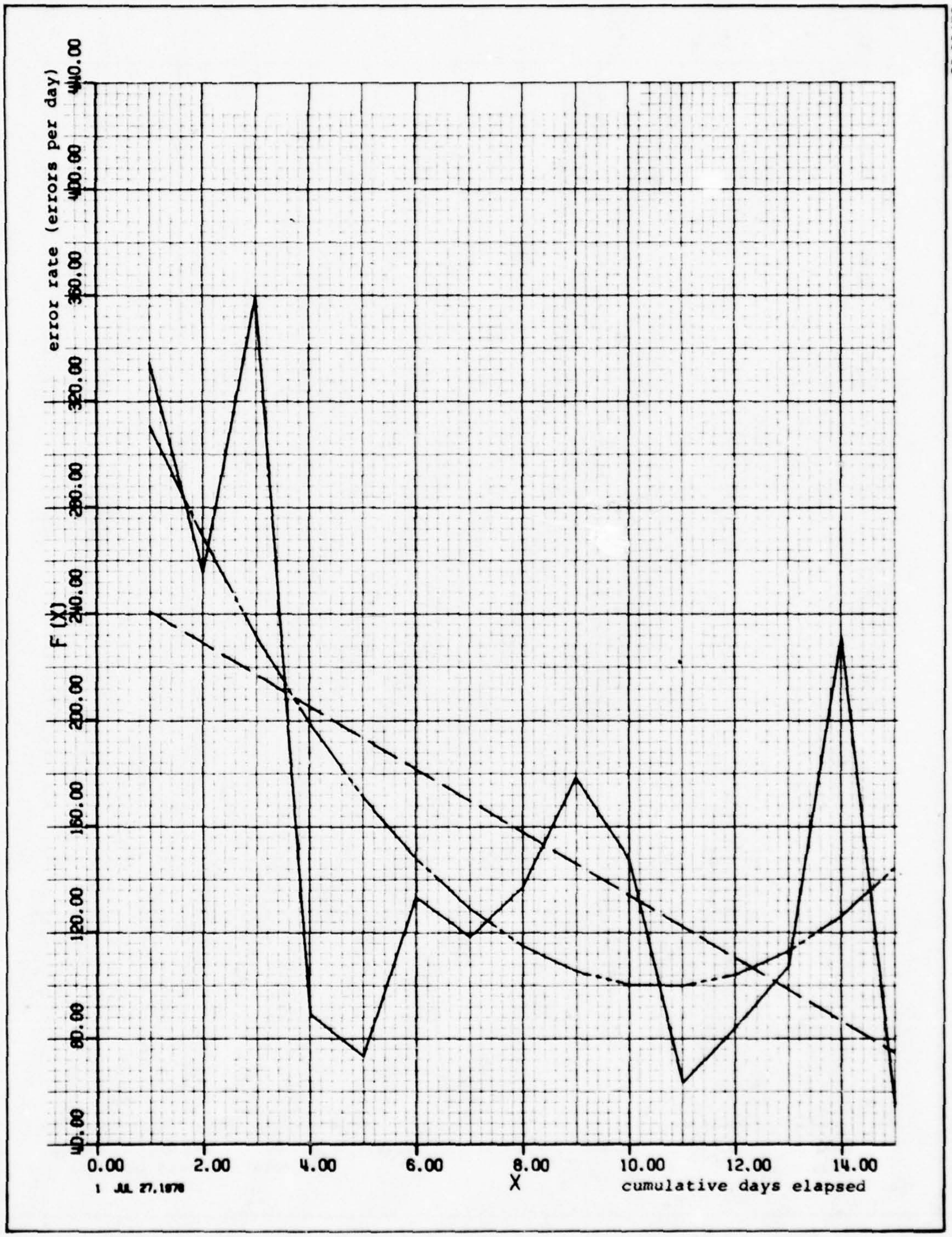

Figure 6.1.16. Error Rate Versus Cumulative Time – Data Set 11-fd

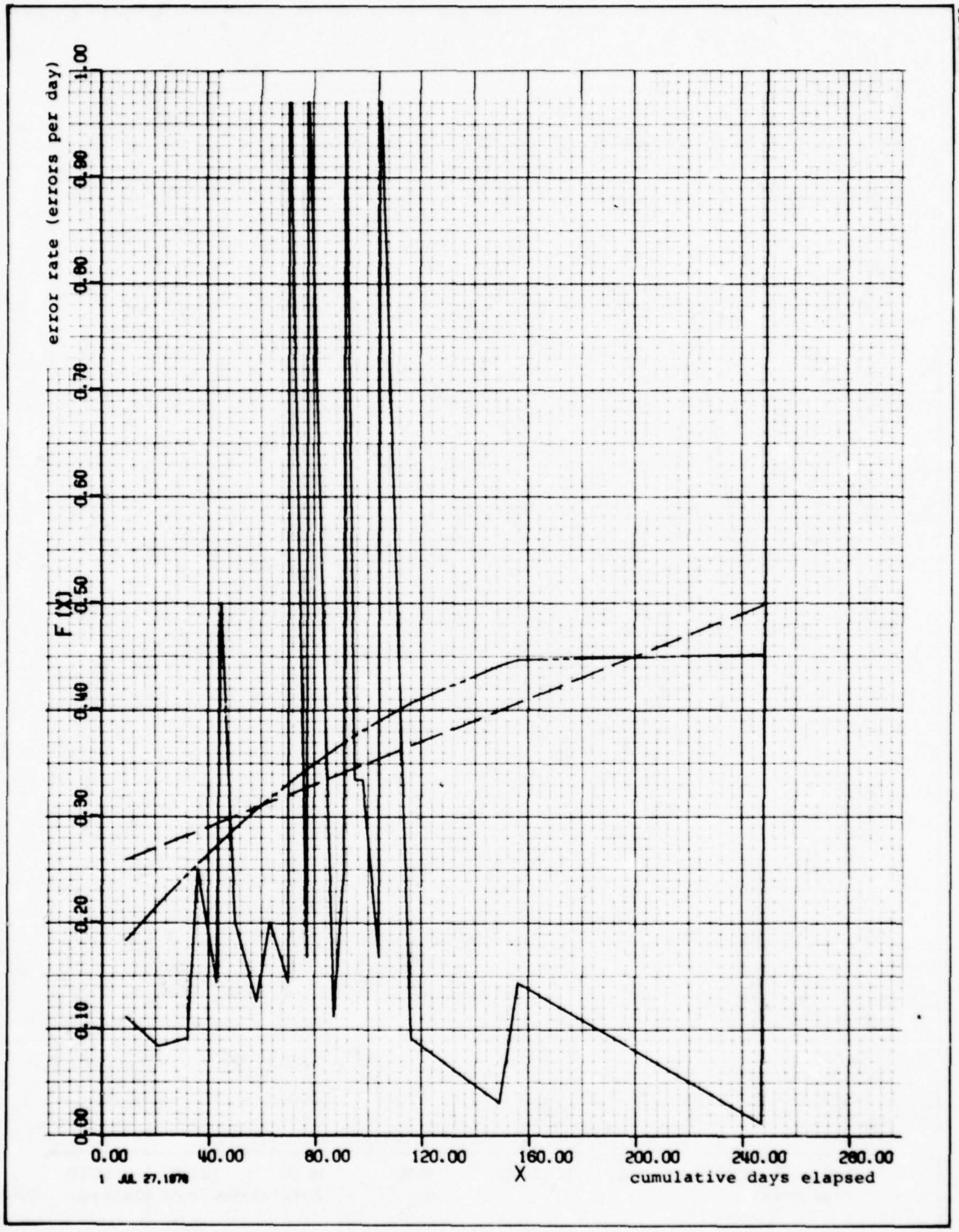

Figure 6.1.17. Error Rate Versus Cumulative Time – Data Set 12-fd

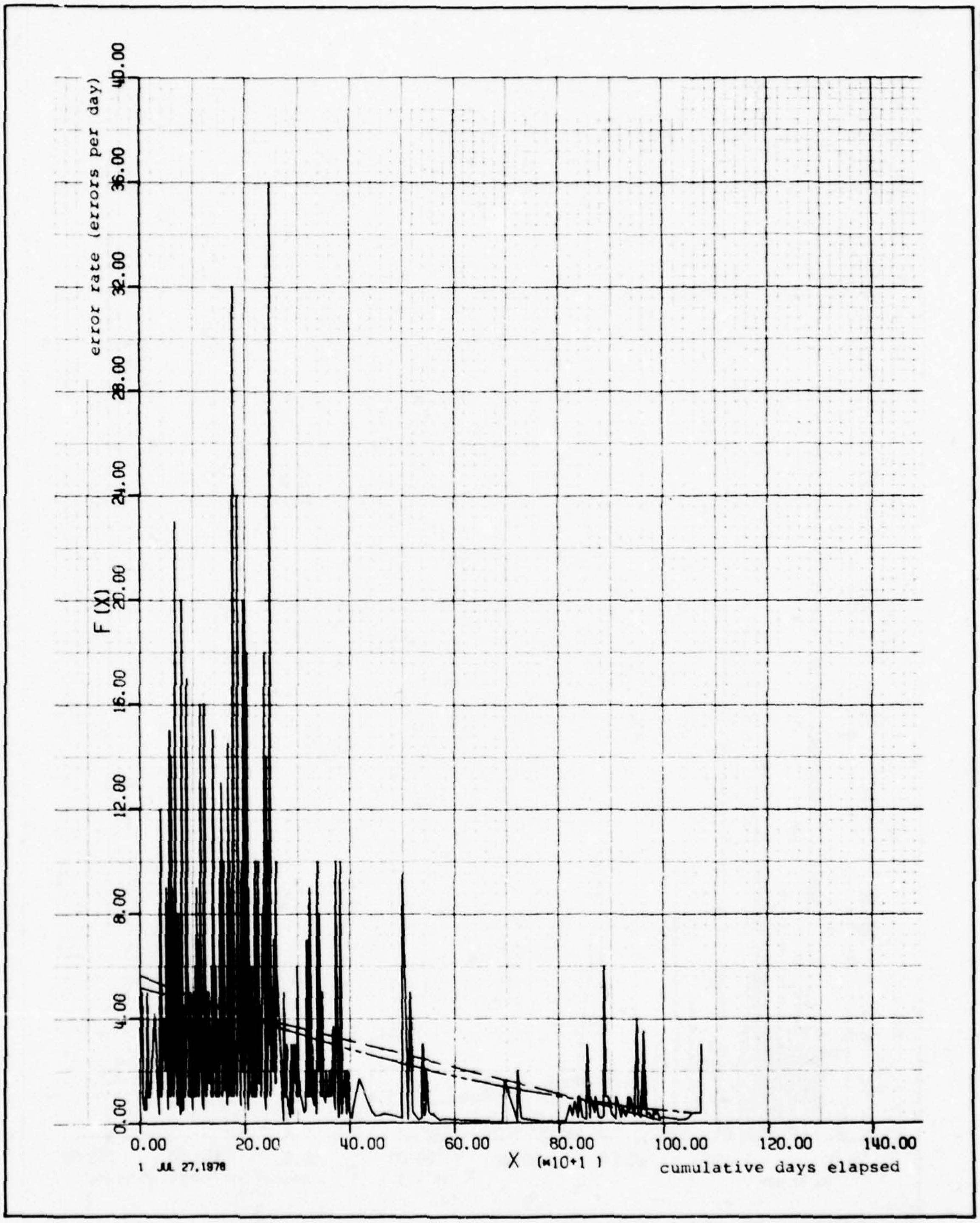

Figure 6.1.18. Error Rate Versus Cumulative Time – Data Set 13-fd

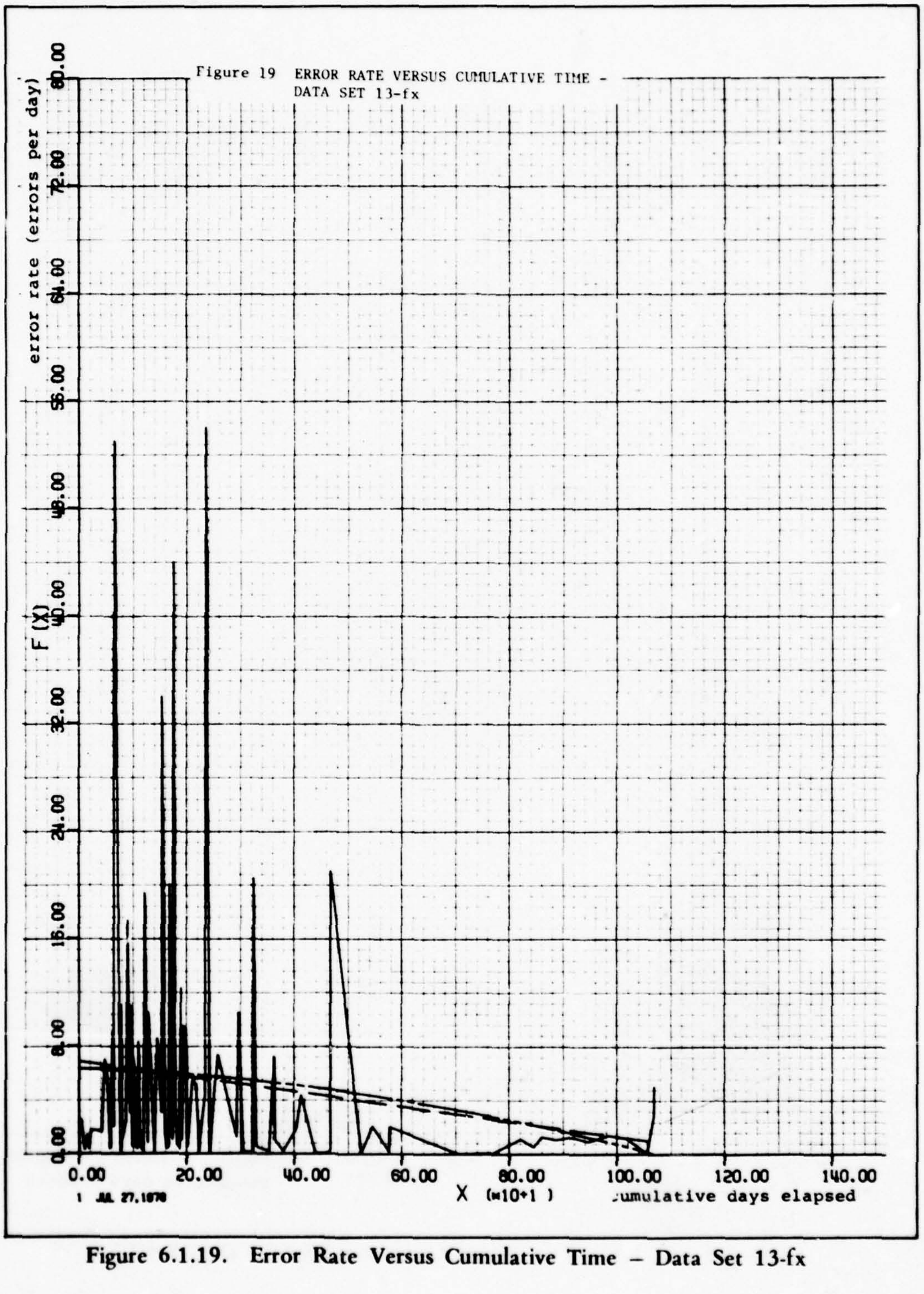

Figure 6.1.19. Error Rate Versus Cumulative Time – Data Set 13-fx

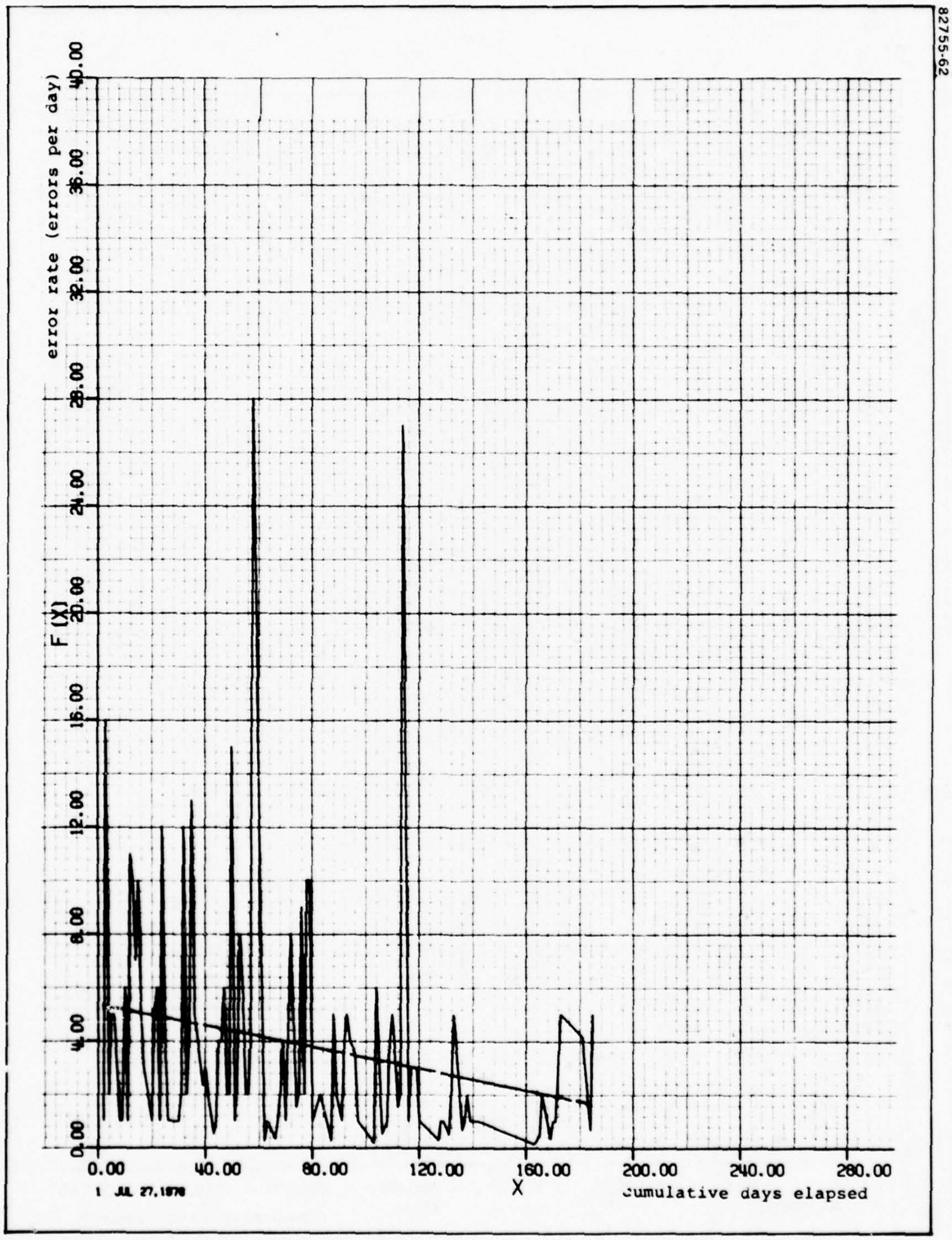

Figure 6.1.20. Error Rate Versus Cumulative Time – Data Set 14-fd

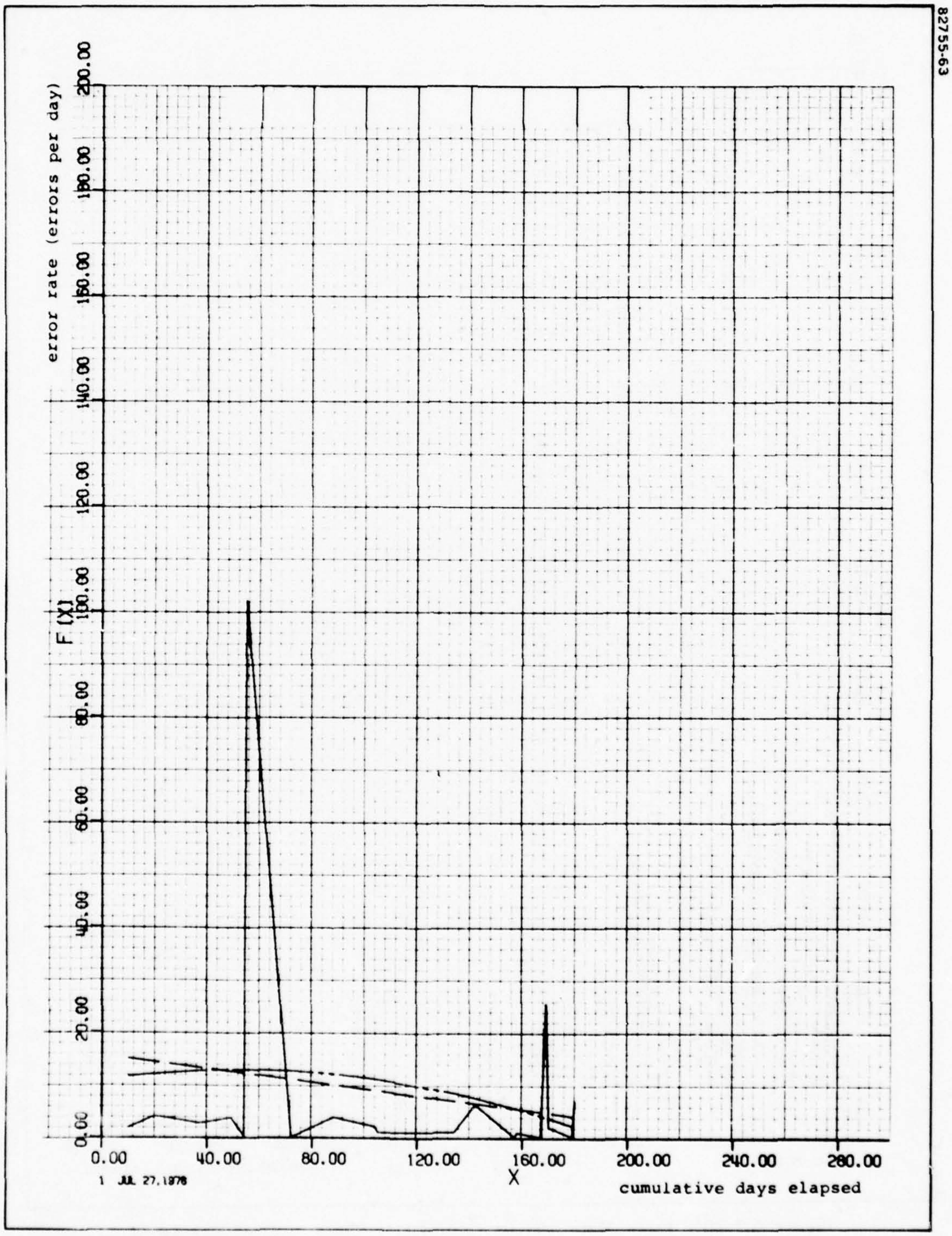

Figure 6.1.21. Error Rate Versus Cumulative Time – Data Set 14-fx

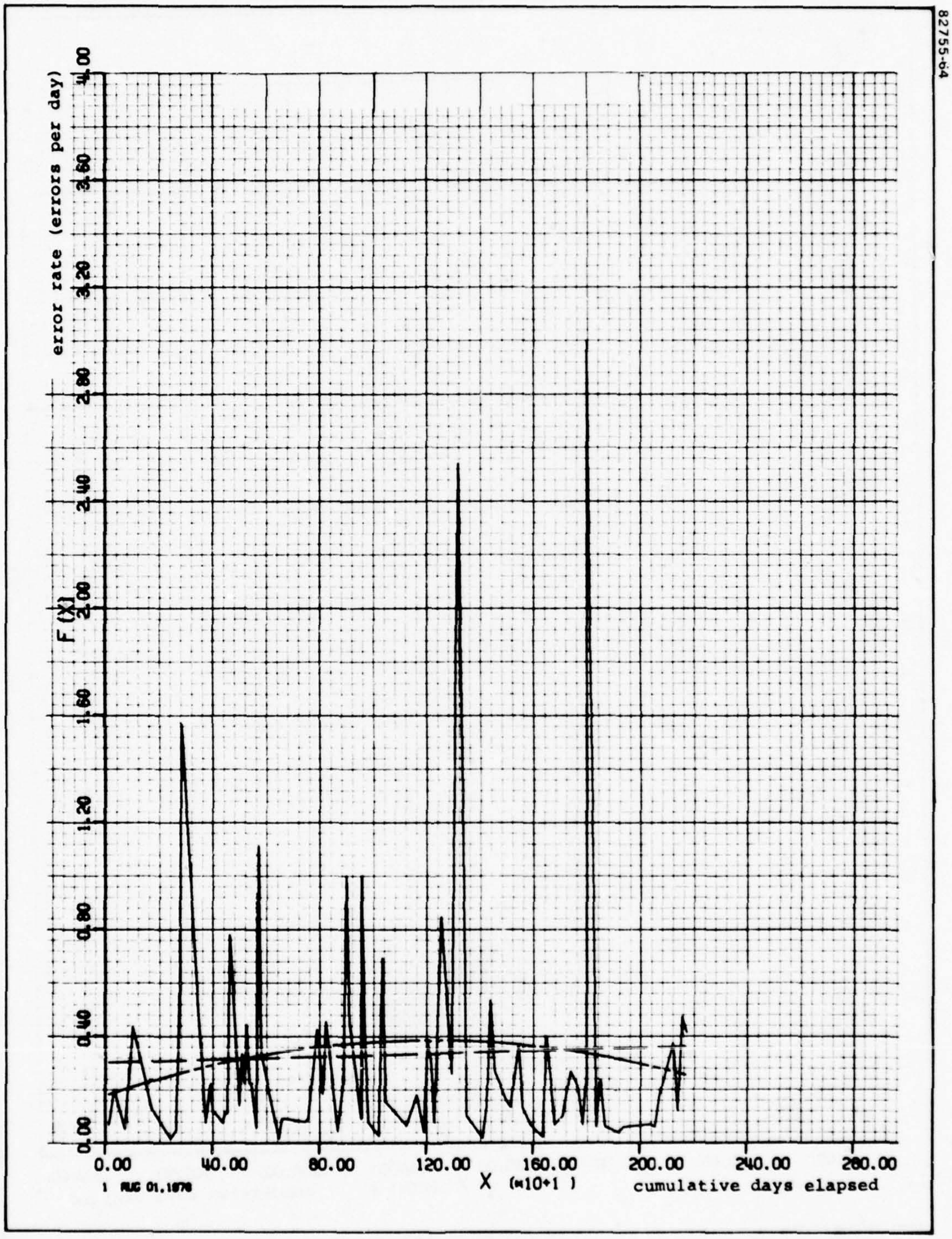

Figure 6.1.22. Error Rate Versus Cumulative Time – Data Set 15-fd

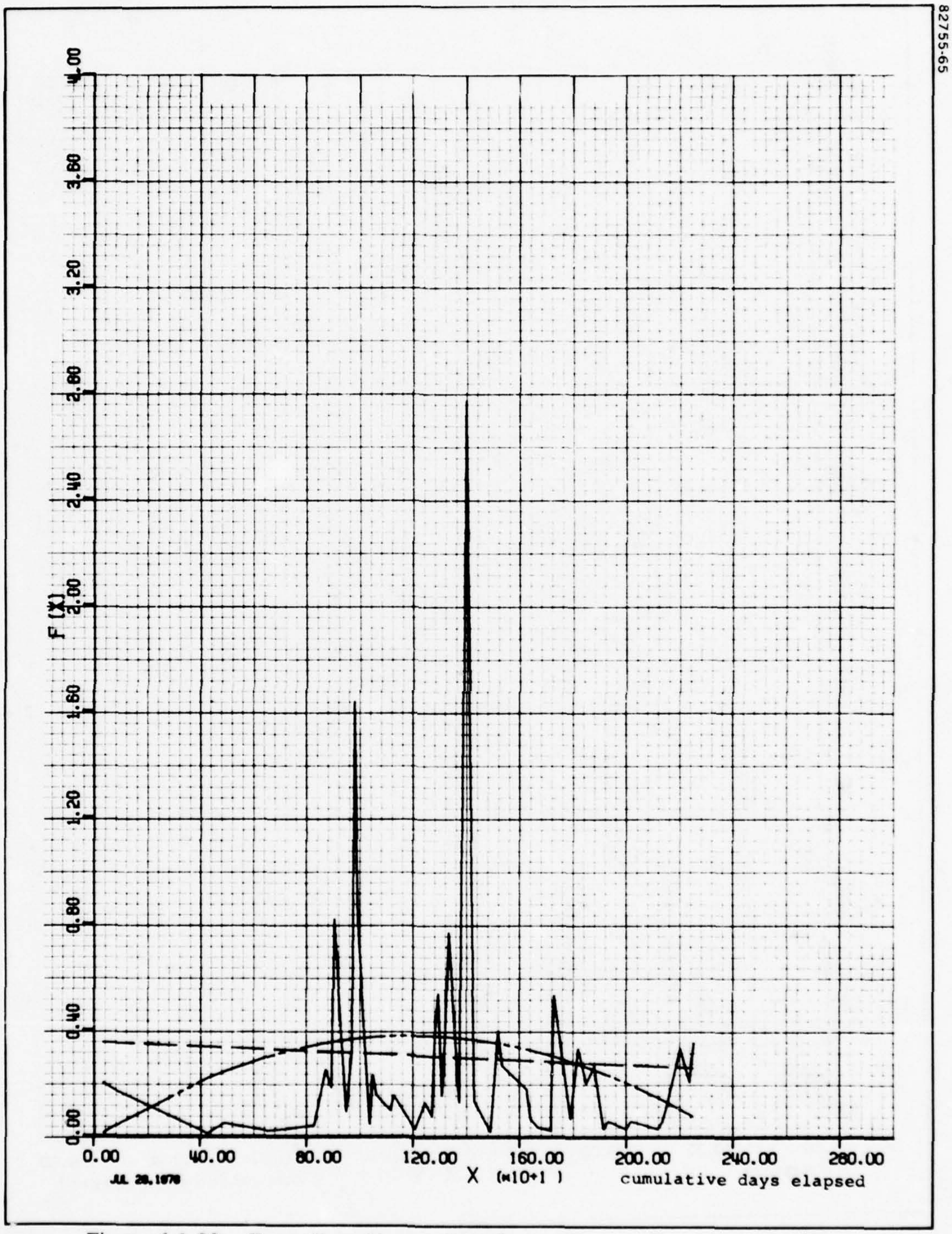

Figure 6.1.23. Error Rate Versus Cumulative Time – Data Set 15A-fd

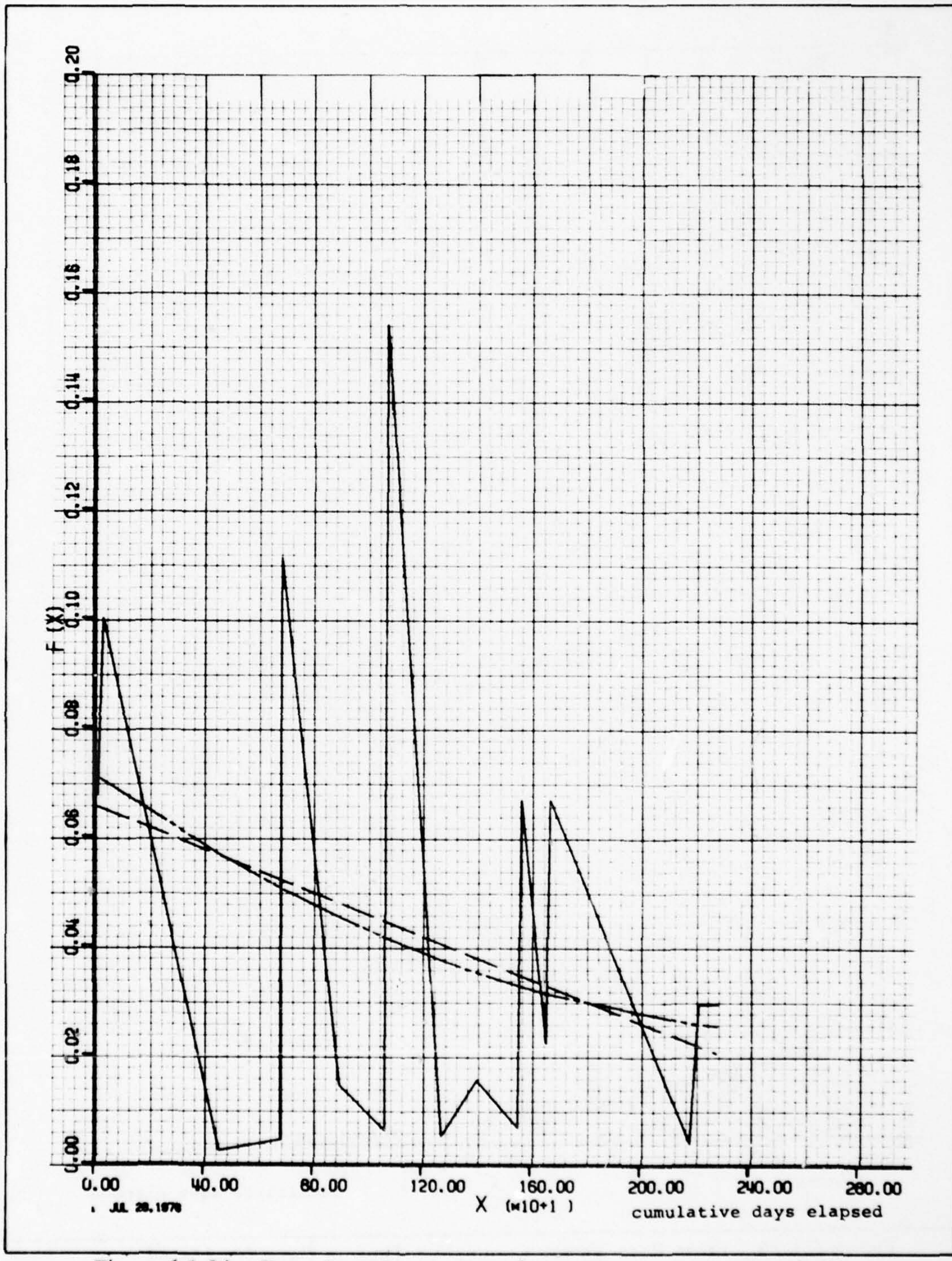

Figure 6.1.24. Error Rate Versus Cumulative Time – Data Set 15B-fd

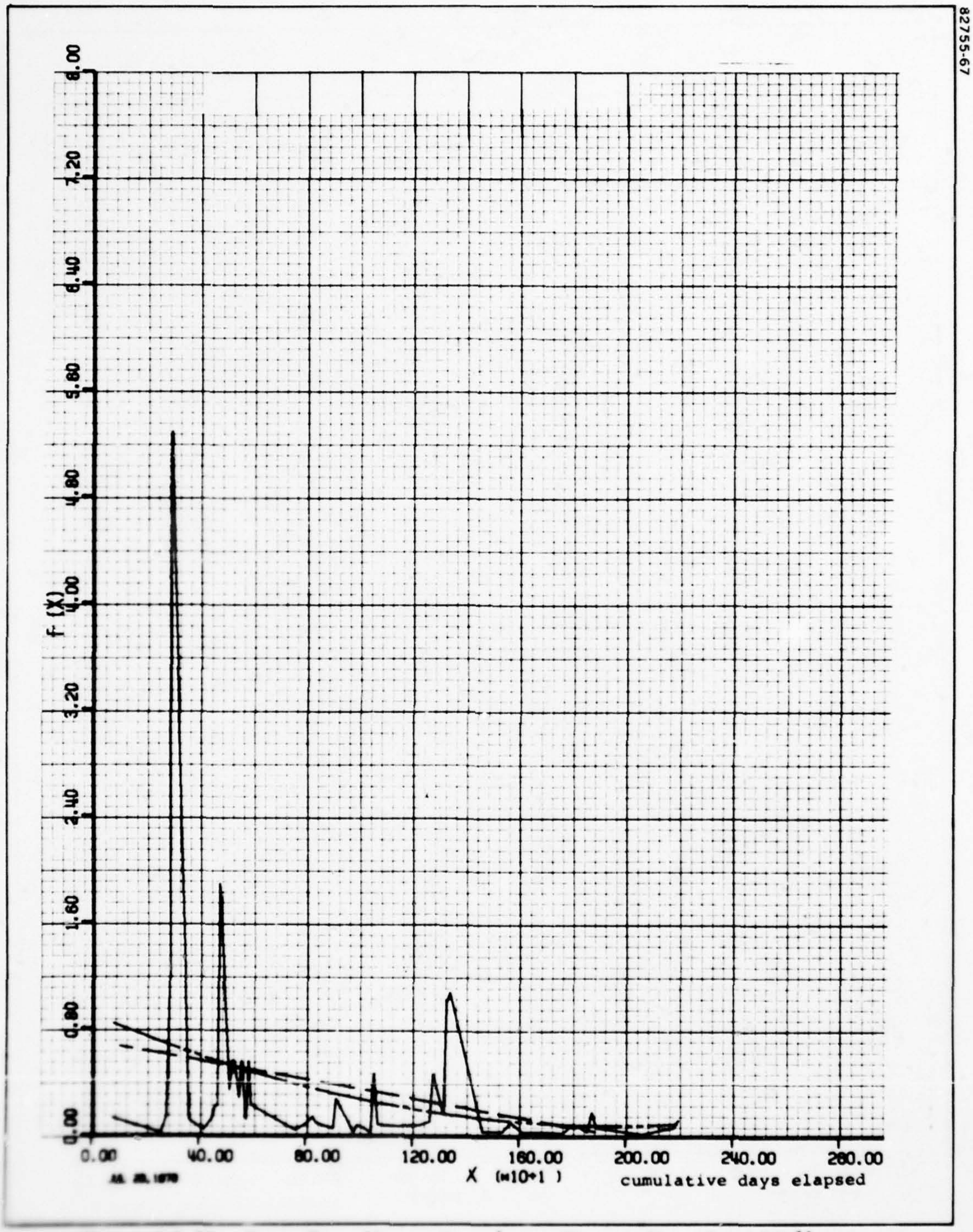

Figure 6.1.25. Error Rate Versus Cumulative Time – Data Set 15C-fd

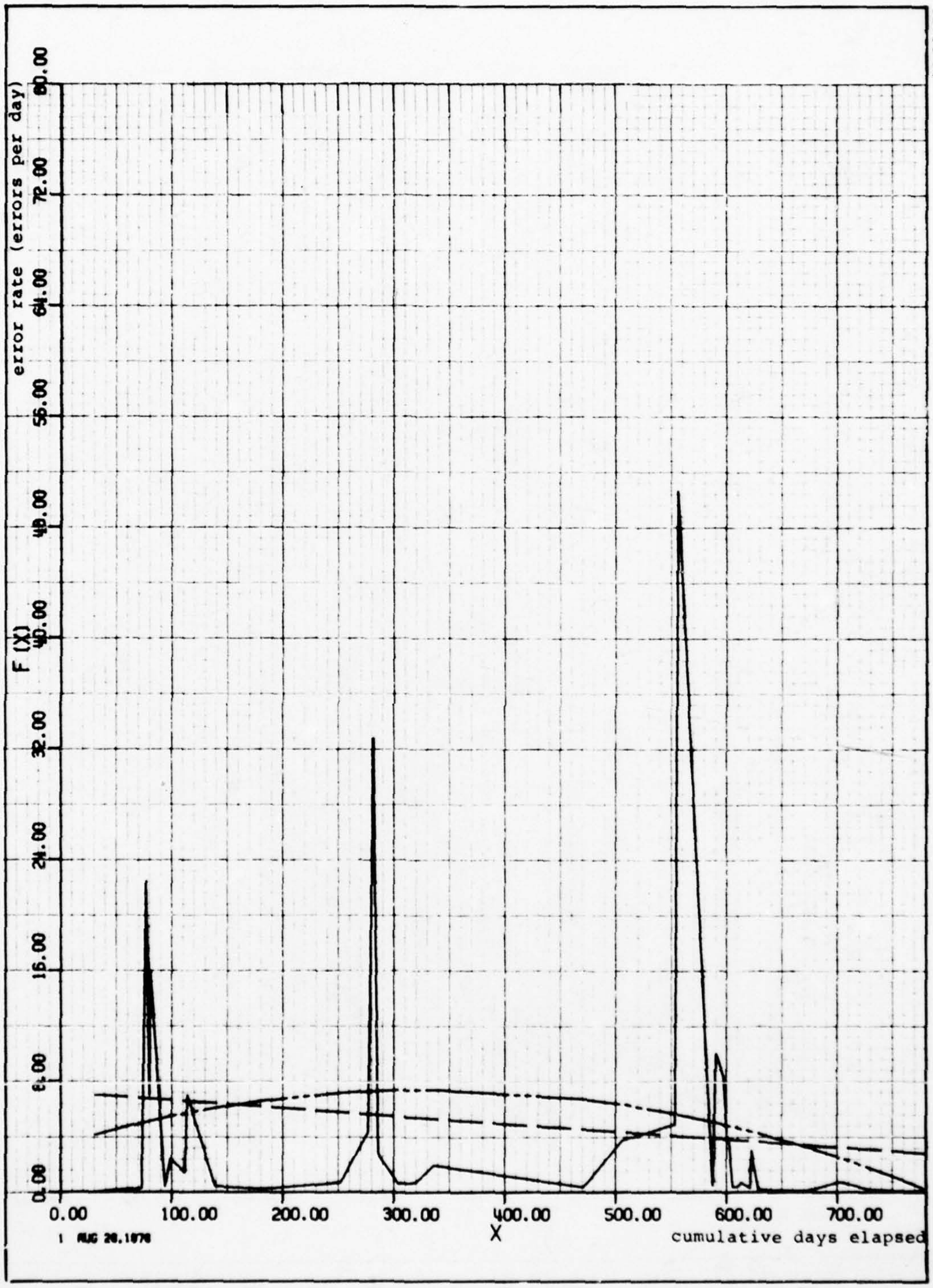

Figure 6.1.26. Error Rate Versus Cumulative Time – Data Set 16-fd

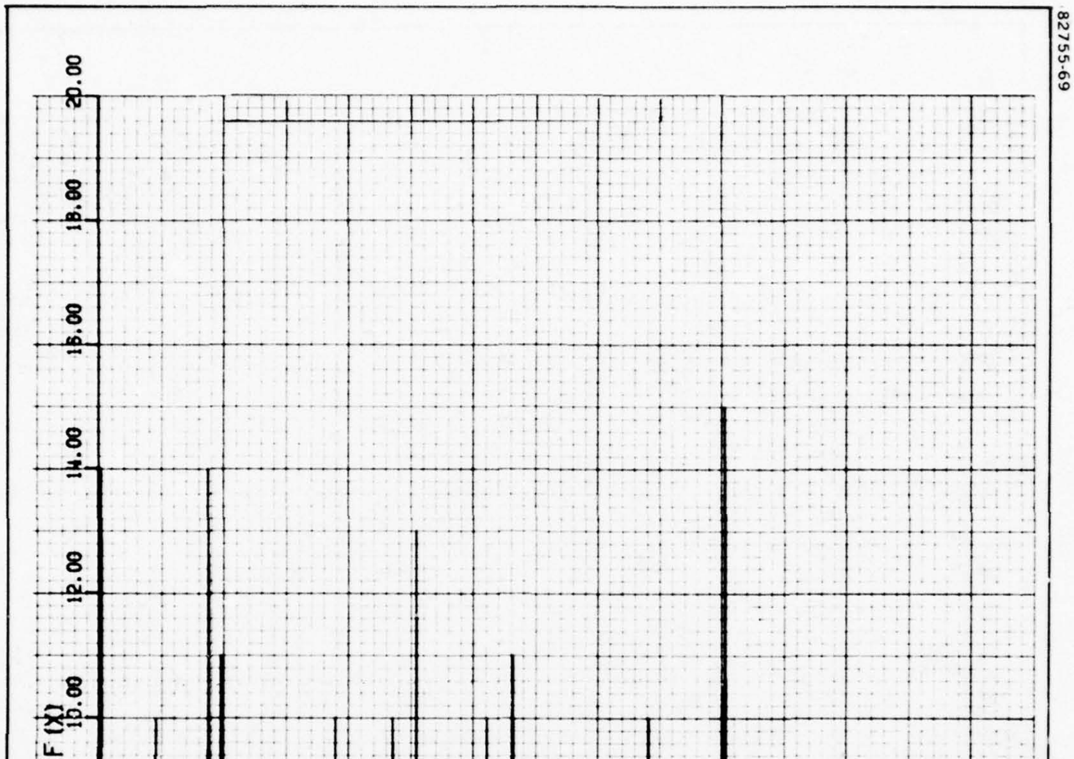

Figure 6.1.27. Error Rate Versus Cumulative Time – Data Set 16-fx

Regarding the assumption of decreasing mean error rate inherent in all of the models, we have two conclusions. First the data sets showed a marked departure from this assumption and the departure appears to be larger than just chance fluctuations. Second the various models are sensitive to this condition.

Another assumption inherent in all of the models is related to the immediately preceding discussion. It is implicit in each model studied (except in the IBM model, in a limited sense) that no errors are introduced in the course of discovering and removing errors. The possibility of introducing errors during the debugging process could also cause the observed increasing error rate condition previously mentioned. Certainly errors are introduced during the debugging process so that this (implicit) assumption of each model is invalid. The magnitude of its effect is unknown to us because we cannot separate the three possible sources of increasing error rate in the data

i) chance fluctuations

ii) increased (as time goes on) manpower/debugging effort

iii) introduction of errors during the debugging process.

For the GPM, S-W and J-M models there is another invalid assumption, which, we must admit, appears to be in the "noise." To assume that the number of errors occurring in some finite interval $\tau_i > 0$ has a Poisson distribution is to assume, implicitly, the possibility of an infinite number of errors in τ_i. Clearly the number of errors must be finite. However if the intervals are reasonably large with a reasonable number of them (intervals) there should be, and we saw, no problem.

Another invalid internal assumption of the Poisson type models is, because the Poisson mean is assumed constant throughout τ_i, the implicit assumption that the errors found in τ_i are removed at the end of the interval. Much the same assumption is true of the binomial model which assumes a constant probability of detecting an error throughout τ_i. We noticed no real effect of this problem.

An excellent feature of the Poisson models is that the (Poisson) joint probability density function of the data (the data are the k pairs (N_i, τ_i)) depends not only on the observed number of errors <u>found</u> in each interval N_i but also the cumulative (to the ith interval) number of errors <u>removed</u> M_{i-1}. Thus, if errors are found but not fixed immediately the Poisson models can accommodate this situation.

Section 7.0

RECOMMENDATIONS FOR FURTHER STUDY AND CONCLUSIONS

The models did not look good in this study - poor fits and lack of convergence of the numerical method. (See Section 0.0 for a summary of the major results of this study and Table 7.2 for a rank-ordering of models according to fit.) Actually little of this was due to poor assumptions inherent (internal) in the models. It is true that the models do not "permit" the introduction of errors in the S/W development process. However, we feel that the failure of the data to show plausible (decreasing error rate, on the average, in increasing time) behavior was not due to the introduction of errors during the debugging process nor to random fluctuations but to the inconsistency of the debugging process. As "proof" of this we note that when one of the models was presumed and a simulation run, the model (from which the data was generated) fit well. Ordinarily this is not enough evidence on which to blame the data. However, the failure of the parameter estimates to converge in so many cases indicates more was going on than just that the models were ill-chosen.

Aside from the problems of poor fits and lack of convergence of the numerical methods, the models (summarized in Table 7.1) are certainly not mathematically unmanageable nor are they totally lacking in physical interpretation in terms of software development phenomena. Indeed, most models contain the parameter N, the total number of errors present in the software initially, about which it is important to make inferences. Also, most of the models were found to be useful, at least in theory, in predicting the time necessary to complete the debugging process. And finally, the parameter estimators were found to be asymptotically normally distributed which, in theory, allows interval estimation of the unknown parameters in large samples using known techniques related to the normal distribution.

Each of the models studied in this report could reasonably serve as an idealized model for large-scale software debugging processes. This, of course, was the intention of the respective "inventors" of the models studied herein. It is obvious, though, that no "idealized" data from large-scale software programs exists and this

problem, until now, has not been recognized. It is conceivable that one or more of these models could describe the actual debugging processes on large-scale software projects very well, or conversely, it may be possible to revise management practices utilized on software projects so that one or more of the models will describe the process. At any rate, we believe that it can be concluded from this study that the state of the art in software data collection practices is entirely incompatible with the sophistication of the models discussed in this report.

Thus we feel there are two areas requiring further study. First, a carefully controlled S/W program shall be designed and the debugging process carefully monitored and error events faithfully recorded. Then, we believe several of the models will be seen to fit. As a part of this (controlled computer program development) study effort standards/methods of S/W error data collection should be developed, the ultimate aim being a MILITARY STANDARD and/or HANDBOOK and/or GUIDELINES for UNIFORM industry S/W error data collection and documentation procedures.

Model (Section)	Mean Values	Joint Probability Function
IBM Model (Sections 3.2, 4.1.5)	$a(1 - e^{-bt}),\ t \geq 0$	$\prod_{i=1}^{k} \frac{\left[a\left(e^{-bt_{i-1}} - e^{-bt_i}\right)\right]^{Z_i}}{Z_i!} e^{-a\left(e^{-bt_{i-1}} - e^{-bt_i}\right)}$
Binomial I (Sections 3,3 4.1.4)	$\left(N - \sum_{j=1}^{i-1} N_j\right) \tau_i(\tau_i + C)^{-1}$	$\frac{N!}{N_1!\, N_2! \cdots N_k! \left(N - \sum_{i=1}^{k} N_i\right)!} \prod_{i=1}^{k} p_i^{N_i}(1 - p_i)^{N - \sum_{j=1}^{i} N_j}$ where $p_i = \tau_i(\tau_i + C)^{-1}$
Binomial II (Section 3,3 4.1.4)	$\left(N - \sum_{j=1}^{i-1} N_j\right)\left(1 - e^{-a\tau_i}\right)$	Same as above with $p_i = \left(1 - e^{-a\tau_i}\right)$

*MLEs not computed

Equations Determining MLEs	Equations Determining LSEs
$\sum_{i=1}^{k} Z_i - \left(1 - e^{-\hat{b}t_k}\right) = 0$	$\sum_{i=1}^{k} Z_i\left(e^{-\tilde{b}t_{i-1}} - e^{-\tilde{b}t_i}\right) - \tilde{a}\sum_{i=1}^{k}\left(e^{-bt_{i-1}} - e^{-bt_i}\right)^2 = 0$
$\sum_{i=1}^{k} \frac{Z_i\left(t_i e^{-\hat{b}t_i} - t_{i-1} e^{-\hat{b}t_{i-1}}\right)}{\left(e^{-\hat{b}t_{i-1}} - e^{-\hat{b}t_i}\right)} - \hat{a}t_k e^{-\hat{b}t_k} = 0$	$\tilde{a}\sum_{i=1}^{k} Z_i\left(t_i e^{-\tilde{b}t_i} - t_{i-1} e^{-\tilde{b}t_{i-1}}\right) - a^2\sum_{i=1}^{k}\left(t_i e^{-\tilde{b}t_i} - t_{i-1} e^{-\tilde{b}t_{i-1}}\right)\left(e^{-\tilde{b}t_{i-1}} -\right.$
*	$\sum_{i=1}^{k} \frac{N_i \tau_i}{(\tau_i + \tilde{C})} - \sum_{i=1}^{k} (\tilde{N} - M_{i-1})\left(\frac{\tau_i}{\tau_i + \tilde{C}}\right)^2 = 0$ $\sum_{i=1}^{k} N_i(\tilde{N} - M_{i-1}) \frac{\tau_i}{(\tau_i + \tilde{C})^2} - \sum_{i=1}^{k} (\tilde{N} - M_{i-1})^2 \frac{\tau_i^2}{(\tau_i + \tilde{C})^3} = 0$
*	$\sum_{i=1}^{k} N_i\left(1 - e^{-\tilde{a}\tau_i}\right) - \sum_{i=1}^{k} (\tilde{N} - M_{i-1})\left(1 - e^{-\tilde{a}\tau_i}\right)^2 = 0$ $\sum_{i=1}^{k} N_i \tau_i(\tilde{N} - M_{i-1})e^{-\tilde{a}\tau_i} - \sum_{i=1}^{k} \tau_i(\tilde{N} - M_{i-1})^2\left(1 - e^{-\tilde{a}\tau_i}\right)e^{-\tilde{a}\tau_i} = 0$

TABLE 7.1. SUMMARY OF MODELS

quations Determining LSEs

$$\sum_{i=1}^{k}\left(e^{-bt_{i-1}} - e^{-bt_{i}}\right)^{2} = 0$$

$$1) - a^{2}\sum_{i=1}^{k}\left(t_{i}e^{-\tilde{b}t_{i}} - t_{i-1}\,e^{-\tilde{b}t_{i-1}}\right)\left(e^{-\tilde{b}t_{i-1}} - e^{-\tilde{b}t_{i}}\right) = 0$$

$$)\left(\frac{\tau_{i}}{\tau_{i} + \tilde{C}}\right)^{2} = 0$$

$$-\sum_{i=1}^{k}(\tilde{N} - M_{i-1})^{2}\;\frac{\tau_{i}^{2}}{\left(\tau_{i} + \tilde{C}\right)^{3}} = 0$$

$$M_{i-1})\left(1 - e^{-\tilde{a}\tau_{i}}\right)^{2} = 0$$

$$\sum_{1}^{k}\tau_{i}(\tilde{N} - M_{i-1})^{2}\left(1 - e^{-\tilde{a}\tau_{i}}\right)e^{-\tilde{a}\tau_{i}} = 0$$

3

Model	Mean Values	Joint Probability Function
Generalized Poisson Model (GPM) (Sections 3.1, 4.1.3)	$\phi(N - M_{i-1})\tau_i^{\alpha}$	$\prod_{i=1}^{k} \frac{\{\phi(N - M_{i-1})\tau_i^{\alpha}\}^{N_i}}{N_i!} e^{-\phi(N - M_{i-1})\tau_i^{\alpha}}$
Jelinski Moranda (Sections 3.1, 4.1.3)	$\phi(N - M_{i-1})\tau_i$	Same as GPM with $\alpha = 1$
Schick-Wolverton (Sections 3.1, 4.1.3)	$\phi(N - M_{i-1})\tau_i^2$	Same as GPM with $\alpha = 2$

TABLE 7.1. SUMMARY OF MODELS-Continued

4

Equations Determining MLEs	Equations Determining LSEs
$\sum_{i=1}^{k} \frac{N_i}{\hat{N} - M_{i-1}} - \hat{\phi} \sum_{i=1}^{k} \tau_i^{\hat{\alpha}} = 0$	$\tilde{\phi}^{2} \sum_{i=1}^{k} (\tilde{N} - M_{i-1}) \tau_i^{2\tilde{\alpha}} - \tilde{\phi} \sum_{i=1}^{k} N_i \tau_i^{\tilde{\alpha}} = 0$
$\sum_{i=1}^{k} N_i - \sum_{i=1}^{k} (\hat{N} - M_{i-1}) \tau_i^{\hat{\alpha}} = 0$	$\tilde{\phi} \sum_{i=1}^{k} (\tilde{N} - M_{i-1})^2 \tau_i^{2\tilde{\alpha}} - \sum_{i=1}^{k} N_i (\tilde{N} - M_{i-1}) \tau_i^{\tilde{\alpha}} = 0$
$\sum_{i=1}^{k} N_i \ln \tau_i - \hat{\phi} \sum_{i=1}^{k} (\hat{N} - M_{i-1}) \tau_i^{\hat{\alpha}} \ln \tau_i = 0$	$\tilde{\phi}^{2} \sum_{i=1}^{k} (\tilde{N} - M_{i-1})^2 \tau_i^{2\tilde{\alpha}} \ln \tau_i - \tilde{\phi} \sum_{i=1}^{k} N_i (\tilde{N} - M_{i-1}) \tau_i^{\tilde{\alpha}} \ln \tau_i = 0$
$\sum_{i=1}^{k} \frac{N_i}{\hat{N} - M_{i-1}} - \hat{\phi} \sum_{i=1}^{k} \tau_i = 0$	$\tilde{\phi}^{2} \sum_{i=1}^{k} (\tilde{N} - M_{i-1}) \tau_i^{2} - \tilde{\phi} \sum_{i=1}^{k} N_i \tau_i = 0$
$\sum_{i=1}^{k} N_i - \sum_{i=1}^{k} (\hat{N} - M_{i-1}) \tau_i = 0$	$\tilde{\phi} \sum_{i=1}^{k} (\tilde{N} - M_{i-1})^2 \tau_i^{2} - \sum_{i=1}^{k} N_i (\tilde{N} - M_{i-1}) \tau_i = 0$
$\sum_{i=1}^{k} \frac{N_i}{\hat{N} - M_{i-1}} - \hat{\phi} \sum_{i=1}^{k} \tau_i^{2} = 0$	$\tilde{\phi}^{2} \sum_{i=1}^{k} (\tilde{N} - M_{i-1}) \tau_i^{4} - \tilde{\phi} \sum_{i=1}^{k} N_i \tau_i^{2} = 0$
$\sum_{i=1}^{k} N_i - \sum_{i=1}^{k} (\hat{N} - M_{i-1}) \tau_i^{2} = 0$	$\tilde{\phi} \sum_{i=1}^{k} (\tilde{N} - M_{i-1})^2 \tau_i^{4} - \sum_{i=1}^{k} N_i (\tilde{N} - M_{i-1}) \tau_i^{2} = 0$

5

1

TABLE 7.2. RANK ORDERING OF MODELS ACCORDING TO FIT

Data Set 1

Model	1-fd	1-fdG	1-fx	1-fxG	1-ff	1-ffG
GPM (LSE)				1		
J-M (LSE)			2	4		
S-W (LSE)			3			
GPM (MLE)				1		
J-M (MLE)						
S-W (MLE)						
Binomial (LSE) I				3		
Binomial (LSE) II						
IBM (LSE)	1	2	2	5		
IBM (MLE)		1	1	2		

Data Set 2

Model	2-fd	2-fdG	2-fx	2-fxG	2-ff	2-ffG
GPM (LSE)				1		
J-M (LSE)			7	6		
S-W (LSE)						
GPM (MLE)			4	1		
J-M (MLE)			1	2		
S-W (MLE)						
Binomial (LSE) I			5	4		
Binomial (LSE) II			6	5		
IBM (LSE)			3	7		
IBM (MLE)			2	3		

Note: Models are ranked from best fit (1) to worst fit (possible 10). Blank entries indicate that no fit was obtained.

TABLE 7.2 RANK ORDERING OF MODELS ACCORDING TO FIT Continued

Data Set 3

Model	3-fd	3-fdG	3-fx	3-fxG	3-ff	3-ffG
GPM (LSE)						
J-M (LSE)						
S-W (LSE)						
GPM (MLE)						
J-M (MLE)						
S-W (MLE)						
Binomial (LSE) I						
Binomial (LSE) II						
IBM (LSE)	1	1	1	1		
IBM (MLE)						

Data Set 4

Model	4-fd	4-fdG	4-fx	4-fxG	4-ff	4-ffG
GPM (LSE)		1			1	1
J-M (LSE)	6	6	6	4	4	3
S-W (LSE)						
GPM (MLE)		1	1		2	1
J-M (MLE)	2	3			3	2
S-W (MLE)						
Binomial (LSE) I	3	4	3	2		
Binomial (LSE) II	5	5	5	3		
IBM (LSE)	4	7	4	5		
IBM (MLE)	1	2	2	1		

Note: Models are ranked from best fit (1) to worst fit (possible 10). Blank entries indicate that no fit was obtained.

TABLE 7.2. RANK ORDERING OF MODELS ACCORDING TO FIT Continued

Data Set 5

Model	5-fd	5-fdG	5-fx	5-fxG	5-ff	5-ffG
GPM (LSE)	2	1	2	1	1	1
J-M (LSE)			7	6		
S-W (LSE)			9	7		
GPM (MLE)	1	1	1	1	1	1
J-M (MLE)			3	2		
S-W (MLE)						
Binomial (LSE) I			5	4		
Binomial (LSE) II			6	5		
IBM (LSE)	3	2	8			
IBM (MLE)			4	3		

Data Set 6

Model	6-fd	6-fdG	6-fx	6-fxG	6-ff	6-ffG
GPM (LSE)						
J-M (LSE)	4	5				
S-W (LSE)	6	7				
GPM (MLE)						
J-M (MLE)						
S-W (MLE)	5	6				
Binomial (LSE) I	3	3				
Binomial (LSE) II	4	4				
IBM (LSE)	2	2				
IBM (MLE)	1	1				

Note: Models are ranked from best fit (1) to worst fit (possible 10). Blank entries indicate that no fit was obtained.

TABLE 7.2. RANK ORDERING OF MODELS ACCORDING TO FIT Continued

Data Set 7

Model	7-fd	7-fdG	7-fx	7-fxG	7-ff	7-ffG
GPM (LSE)	1	1	2	2		1
J-M (LSE)	5	5	8	7		
S-W (LSE)	8	8	10	10		
GPM (MLE)	1	1	1	1		1
J-M (MLE)	2	2	4	3		
S-W (MLE)	7	7	9	9		
Binomial (LSE) I	4	4	6	5		
Binomial (LSE) II	4	5	7	6		
IBM (LSE)	6	6	5	8		
IBM (MLE)	3	3	3	4		

Data Set 8

Model	8-fd	8-fdG	8-fx	8-fxG	8-ff	8-ffG
GPM (LSE)						
J-M (LSE)	3					
S-W (LSE)	3					
GPM (MLE)						
J-M (MLE)	1					
S-W (MLE)	1					
Binomial (LSE) I	3					
Binomial (LSE) II	3					
IBM (LSE)	4					
IBM (MLE)	2					

Note: Models are ranked from best fit (1) to worst fit (possible 10). Blank entries indicate that no fit was obtained.

TABLE 7.2. RANK ORDERING OF MODELS ACCORDING TO FIT Continued

Data Set 9

Model	9-fd	9-fdG	9-fx	9-fxG	9-ff	9-ffG
GPM (LSE)						
J-M (LSE)	3					
S-W (LSE)	3					
GPM (MLE)						
J-M (MLE)	2					
S-W (MLE)	2					
Binomial (LSE) I						
Binomial (LSE) II	3					
IBM (LSE)	2					
IBM (MLE)	1					

Data Set 10

Model	10-fd	10-fdG	10-fx	10-fxG	10-ff	10-ffG
GPM (LSE)						
J-M (LSE)	2					
S-W (LSE)	2					
GPM (MLE)						
J-M (MLE)	1					
S-W (MLE)	1					
Binomial (LSE) I	2					
Binomial (LSE) II	2					
IBM (LSE)	4					
IBM (MLE)	3					

Note: Models are ranked from best fit (1) to worst fit (possible 10). Blank entries indicate that no fit was obtained.

GLOSSARY

LSE	–	least squares estimate
LS	–	least squares
MLE	–	maximum likelihood estimate
ML	–	maximum likelihood
iid	–	independent and identically distributed
S/W	–	software
SPR	–	Software Problem Report
PTR	–	Program Trouble Report (Hughes version of SPR)
$\hat{\cdot}$	–	MLE of parameter $\cdot$
$\tilde{\cdot}$	–	LSE of parameter $\cdot$
J-M	–	Jelinski-Moranda
S-W	–	Schick-Wolverton
Var $(\cdot)$	–	Variance of $\cdot$
E $(\cdot)$	–	Expectation of $\cdot$
Cov $(\cdot, \cdot\cdot)$	–	Covariance of $\cdot$ and $\cdot\cdot$
GPM	–	Generalized Poisson Model
N	–	Total number of errors initially present in S/W
N_i	–	The number errors observed in the ith debugging-time interval (not to be confused with N).
k	–	number of debugging-time intervals
$\overset{a}{=}$	–	asymptotically equal to (in probability when random quantities are compared)
N (0, 1)	–	Normally distributed with mean 0 and variance 1.

Section 9.0

BIBLIOGRAPHY

I. Books

[1] Buck, Robert Creighton
Advanced Calculus
McGraw-Hill Book Co., Inc.
New York, 1956

[2] Loeve, Michel Moise
Probability Theory
D. Van Nostrand Co., Inc.
New York, 1963
Third Edition

[3] Rao, C. Radhakrishna
Linear Statistical Inference and Its Applications, 2nd ed.
John Wiley and Sons
New York, 1973

[4] Record 1973 IEEE Symposium on Computer Software Reliability
Institute of Electrical and Electronics Engineers, Inc.
New York, 1973

[5] Seale, S. R.
Linear Models
John Wiley and Sons
New York, 1971

II. Articles

[1] Chernoff, H. and Lehmann, E. L.
"The Use of the Maximum Likelihood Estimates in X^2 Tests for Goodness of Fit"
Ann. Math. Statist., 1954, 25, pp. 579-586

[2] Goel, Amrit L. and Okumoto, K.
Software Failure Analysis by Non-Homogeneous Poisson Process, unpublished
Syracuse University
Syracuse, New York

[3] Lipow, Myron
Software Reliability
Presented at the 12th Annual Reliability Engineering and Management Institute
The University of Arizona, Tucson, November 1974

[4] Molinari, Luciano
"Distribution of the Chi-Squared Test in Nonstandard Situations"
Biometrika, 1977, 64, 1, pp. 115-121

[5] Moran, P. A. P.
"A Mathematical Theory of Animal Trapping"
Biometrika, 1951, (38), pp. 307-311

[6] Moranda, P. B.
"Prediction of Software Reliability During Debugging"
Proceedings 1975 Annual Reliability and Maintainability Symposium,
pp. 327-332

[7] Nagle, H. T., Jr.
Applicability of Reliability Techniques to BMD Software
Auburn University, Engineering Experiment Station
Auburn, Alabama, October 15, 1975

[8] Watson, G. S.
"On Chi-Square Goodness-of-fit Tests for Continuous Distributions"
J. R. Statist. Soc., 1958, B20, pp. 44-61

III. Technical Reports

[1] Baker, W. F.
Software Data Collection and Analysis, RADC-TR-77-192, Technical Report
Rome Air Development Center, Air Force Systems Command
Griffiss Air Force Base, New York, January 1977, AD# A041 644

[2] Brooks, W. D. and Weiler, P. W.
Software Reliability Analysis, IBM Technical Report FSD 77-0009
International Business Machines Corporation, Federal Systems Division
Gaithersburg, Maryland

[3] Doty, D. L., P. J. Nelson, and K. R. Stewart
Software Cost Estimation Study, Guidelines for Improved Software Cost Estimating, RADC-TR-77-220, Vol. II (of two), Final Technical Report
Rome Air Development Center, Air Force Systems Command
Griffiss Air Force Base, New York, August 1977, AD# A044 609

[4] Fries, M. J.
Software Error Data Acquisition, RADC-TR-77-130, Technical Report
Rome Air Development Center, Air Force Systems Command
Griffiss Air Force Base, New York, November 1976, AD# A039 916

[5] Goel, Amrit L. and Okumoto, K.
Bayesian Software Prediction Models, Bayesian Software Correction Limit Policies, RADC-TR-78-155, Vol. IV (of five), Final Technical Report
Rome Air Development Center, Air Force Systems Command
Griffiss Air Force Base, New York, July 1978, AD# A057 873

[6] House, Carol C.
A Within Year Growth Model Approach to Forecasting Corn Yields
Economics, Statistics, and Cooperatives Service, U. S. Department of Agriculture, Crop Reporting Board
Washington, D. C., September 1977

[7] Laemmel, A. and Shooman, M.
Software Modeling Studies, Statistical (Natural) Language Theory and Computer Program Complexity, RADC-TR-78-4, Vol. II (of two), Final Technical Report
Rome Air Development Center, Air Force Systems Command
Griffiss Air Force Base, New York, April 1978, AD# A054 900

[8] Larsen, Greg A.
Alternative Methods of Adjusting for Heteroscedasticity in Wheat Growth Data
Economics, Statistics, and Cooperatives Service, U. S. Department of Agriculture, Crop Reporting Board
Washington, D. C., February 1978

[9] Lipow, Myron
Maximum Likelihood Estimation of Parameters of a Software Time-to-Failure Distribution, TRW Report No. 2260.1.9-73D-15 (Rev 1)
TRW Systems Group
Redondo Beach, California, June 1973

[10] Moranda, P. B. and Jelinski, Z.
Final Report on Software Reliability Study, MDC Report No. 63921
McDonnell Douglas Astronautics Company
Advance Information Systems, December 1972

[11] Motley, R. W. and Brooks, W. D.
Statistical Prediction of Programming Errors, RADC-TR-77-175, Final Technical Report, AD# A041 106
Rome Air Development Center, Air Force Systems Command
Griffiss Air Force Base, New York, May 1977

[12] Rudner, B.
Seeding/Tagging Estimation of Software Errors: Models and Estimates, RADC-TR-77-15, Technical Report, AD# A036 655
Rome Air Development Center, Air Force Systems Command
Griffiss Air Force Base, New York, January 1977

[13] Rye, P., et al
Software Systems Development: A CDSL Project History
RADC-TR-77-213, Technical Report, AD# A042 186
Rome Air Development Center, Air Force Systems Command
Griffiss Air Force Base, New York, June 1977

[14] Shooman, M. and Natarajan, S.
Effect of Manpower Deployment and Bug Generation on Software Error Models, RADC-TR-76-400, Technical Report, AD# A036 106
Rome Air Development Center, Air Force Systems Command
Griffiss Air Force Base, New York, January 1977

[15] Shooman, M. L. and Ruston, H.
Software Modeling Studies, Summary of Technical Progress, RADC-TR-78-4, Vol. I (of two), Final Technical Report, AD# A052 615
Rome Air Development Center, Air Force Systems Command
Griffiss Air Force Base, New York, January 1978

[16] Shooman, M. L. and Ruston H.
Summary of Technical Progress, Software Modeling Studies, RADC-TR-76-143, Technical Report, AD# A025 895
Rome Air Development Center, Air Force Systems Command
Griffiss Air Force Base, New York, May 1976

[17] Shooman, M. L. and Ruston, H.
Summary of Technical Progress, Software Modeling Studies, RADC-TR-76-405, Technical Report, AD# A036 721
Rome Air Development Center, Air Force Systems Command
Griffiss Air Force Base, New York, January 1977

[18] Shooman, M. L. and Ruston, H.
Summary of Technical Progress, Software Modeling Studies, RADC-TR-77-88, Technical Report, AD# A038 508
Rome Air Development Center, Air Force Systems Command
Griffiss Air Force Base, New York, March 1977

[19] Stover, Robert E., Jr.
A Statistics Collection Package for the Jovial J3 Programming Language, RADC-TR-77-293, In-House Report AD# A044 947
Rome Air Development Center, Air Force Systems Command
Griffiss Air Force Base, New York, September 1977

[20] Sukert, Captain Alan N.
A Software Reliability Modeling Study, RADC-TR-76-247, In-House Report
Rome Air Development Center, Air Force Systems Command, AD# A030 437
Griffiss Air Force Base, New York, August 1976

[21] Thayer, T. A., et al
Software Reliability Study, RADC-TR-76-238, Final Technical Report
Rome Air Development Center, Air Force Systems Command, AD# A030 798
Griffiss Air Force Base, New York, August 1976

[22] Trivedi, Ashok K. and Shooman, Martin L.
A Markov Model for the Evaluation of Computer Software Performance,
PI-EEIEP 74-011-EER110
Polytechnic Institute of New York, Department of Electrical Engineering
and Electrophysics
Brooklyn, New York, May 1, 1974

[23] William, H. E., et al
Software Systems Reliability: A Raytheon Project History,
RADC-TR-77-188, Final Technical Report, AD# A040 992
Rome Air Development Center, Air Force Systems Command
Griffiss Air Force Base, New York, June 1977

[24] Wolverton, R. W. and Schick, G. J.
Assessment of Software Reliability, TRW-55-72-04
TRW Systems Engineering and Integration Division
Redondo Beach, California, September 1972

☆U.S. GOVERNMENT PRINTING OFFICE:1979-614-023/187

MISSION
of
Rome Air Development Center

RADC plans and executes research, development, test and selected acquisition programs in support of Command, Control Communications and Intelligence (C^3I) activities. Technical and engineering support within areas of technical competence is provided to ESD Program Offices (POs) and other ESD elements. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.